Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Stanford University*
Donald B. Lindsley, *Northwestern University*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Margaret K. Harlow, Managing Editor

CONTENTS

Conditioned Fear as Revealed by Magnitude of Startle Response to an Auditory Stimulus: J. S. BROWN, H. I. KALISH, AND I. E. FARBER 317

The Intelligibility of Speech as a Function of the Context of the Test Materials: G. A. MILLER, G. A. HEISE, AND W. LICHTEN 329

Sensory Cues in Pitch Judgment: L. M. BRAMMER 336

Behavior Strength as a Function of the Intensity of the Hunger Drive: G. A. KIMBLE 341

A Verification of the Guessing-Sequence Hypothesis about Spread of Effect: C. A. FAGAN AND A. J. NORTH 349

Conditioning Imagery: C. LEUBA AND R. DUNLAP 352

Facilitation and Interference in Performance on the Modified Mashburn Apparatus: II. The Effects of Varying the Amount of Interpolated Learning: DOROTHY E. MCALLISTER AND D. LEWIS 356

Two-Point Discrimination in Visual Space as a Function of the Temporal Interval between the Stimuli: M. LEYZOREK 364

The Influence of Figure-Ground Relationships in Binocular Rivalry: L. T. ALEXANDER 376

The Effect of Differential Nonreinforcement of the Incorrect Response on the Learning of the Correct Response in the Simple T-*Maze:* M. R. DENNY AND M. D. DUNHAM 382

American Psychological Association

Vol. 41 No. 5 May 1951

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., Prince and Lemon Sts., Lancaster, Pa., or 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 25 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 41 NO. 5 MAY 1951

CONDITIONED FEAR AS REVEALED BY MAGNITUDE OF STARTLE RESPONSE TO AN AUDITORY STIMULUS [1]

BY JUDSON S. BROWN, HARRY I. KALISH, AND I. E. FARBER

State University of Iowa

During recent years considerable emphasis has been placed upon the importance of fear or anxiety as a determinant of behavior. According to one interpretation (**12, 13**), fear is a conditioned, anticipatory response to impending painful stimulation. This response, it is further assumed, functions as a secondary motivational system and should exhibit, therefore, certain of the major properties commonly ascribed to primary drive states. In particular, (1) fear should act to intensify other response tendencies during the period of its evocation, and (2) it should, if reduced in strength following a response, function as a reinforcing event leading to the strengthening of that response.

In experimental studies designed to explore this conception, the resulting conclusions have been based almost entirely upon demonstrations of the reinforcing properties of fear reduction (**2, 4, 9, 11, 14**). Typically, the procedures have been of the following general sort. A period of training is first introduced during which a neutral stimulus (or stimuli) is paired with a noxious one (usually electric shock). The noxious stimulus is presumed to act as a UCS eliciting a pain reaction, and the neutral one is assumed to become the CS. Following the initial training trials, the animal is presented with the CS alone. On the basis of the foregoing assumptions, any response that occurs during the presentation of the CS should be strengthened if it is followed closely by the elimination or diminution of the conditioned (fear-arousing) stimulus. The presence of fear under these conditions is demonstrated by the fact that new responses are learned, and by the fact that old responses (those occurring during the conditioning trials) may show extreme persistence (**4**) despite the absence of primary reinforcement.

Although the successful clarification of a number of problems has resulted from the use of this fear-reduction method, it suffers from several practical disadvantages. For instance, it is impossible by means of this procedure to determine the latency with which the fear response is aroused following the onset of the CS, or to determine the course of its development

[1] This study was supported in part by a research grant from the United States Public Health Service covering Research Project MH-142C, entitled *Anxiety and Frustration in Human and Animal Behavior.*

during the time the CS is acting. Furthermore, the testing trials, during which the UCS is omitted, may constitute extinction trials insofar as fear is concerned. If the response to be learned is low in initial relative strength, the fear may extinguish before the reaction has occurred frequently enough to become learned. And finally, if the conditions of the testing situation differ markedly from those obtaining during the conditioning trials, the animals may learn rather quickly to differentiate between the conditioning and test trials, i.e., fear may not be elicited in the test situation. This could occur either because the cues present during testing are inadequate to arouse fear or because they elicit responses that inhibit fear.

In view of these limitations of the fear-reduction method, it has appeared desirable to explore other procedures. The present paper reports one such method that appears to hold considerable promise for future investigations in the area of fear in particular and in the field of conditioning in general. This method is based upon the presumed intensifying or motivating properties of fear rather than upon its reinforcing properties.

On the assumption that fear has drive properties, it follows within the framework of Hull's theory (**5**) that stimulus-response tendencies should be enhanced by its presence. It is immaterial whether such tendencies are the result of previous learning or whether they are innate. In theory, if the response being measured is highest in the animal's repertoire (including responses that may accompany fear), that response will occur more frequently and will have a greater amplitude and shorter latency when fear is present than when it is not. Furthermore, the stronger the fear, the more marked should be its effect upon these measurable features of the response.[2]

In the present study, the response selected for measurement was the startle reaction of rats to an explosive auditory stimulus. One of the original bases for choosing the startle response was the frequently heard clinical observation that anxious patients exhibit exaggerated startle reactions to sudden loud sounds. It does not follow from this alone, of course, that rats will necessarily behave like clinical patients; but if fear has motivational properties in patients as well as in rats, the theory would predict that both would show magnified startle responses when anxious.

In general, the present method involved the use of a stabilimeter-like apparatus capable of recording the magnitude of a rat's jump to a loud sound. In order to demonstrate the utility of the technique, rats were given a series of trials designed to condition the fear reaction, with interspersed test trials to measure the startle response. The training trials were followed by three series of extinction trials. As will be shown below, the startle response was found to increase as a function of the number of fear-conditioning trials, to undergo progressive diminution (extinction) when

[2] It should be noted that these statements are not inconsistent with the observation that an increase in drive does not lead to the indiscriminate strengthening of *every* response. One must consider the nature of the responses that are elicited by the stimuli associated with the drive itself. For instance, Estes and Skinner (**3**) have shown that the presence of fear may inhibit bar-pressing, and Miller, Brown, and Lipofsky (cited in **10**) have shown that under proper conditions, increasing fear strengthens overt avoidance but not approach, whereas increasing hunger strengthens approach but not avoidance. It is only when the drive stimuli do not themselves elicit competing response tendencies that an increase in drive will lead to an intensification of the overt responses ordinarily elicited in a given situation.

the UCS was omitted, and to exhibit spontaneous recovery from extinction after an interval of 24 hr.

Apparatus

The apparatus used in recording the startle reactions of the rats was a modified version of a rat stabilimeter previously employed by one of the present authors (1). A schematic drawing of this device is reproduced in Fig. 1. The basic sensing element was a postage scale (Hanson, Model 1509, 5-lb. capacity) from which the platform, the calibrated dial with its protective glass, and the pointer had been removed. A new platform, 2.75 in. wide and 7.5 in. long, made of .25-in. plywood, was attached to the scale. A grid, constructed of brass welding rods (.094-in. diameter) spaced at .437-in. intervals in bakelite supporting strips, was secured to the platform. A space of .5 in. was provided under the grid to reduce the possibility of its being shorted by feces or urine.

Surmounting the grid was a detachable, bottomless confinement box (2.37 in. wide, 7.37 in. long, 5.5 in. high, inside dimensions). The top and front side were formed of thin Plexiglass, and the two ends and the other side were of thin plywood. The box, the mounting platform with its supporting rod, and the grid weighed a total of 487 gm. When a rat was placed on the grid and the confinement box fastened down around it, any loud sharp sound would produce a startle reaction that would cause the platform to be momentarily depressed below its normal position. These downward movements of the platform were transmitted by means of a thread to an ink-writing muscle lever whose deflections were recorded on a 5-in. tape moving at .53 in. per sec. Prolonged oscillations of the system were prevented by a damping device consisting of a plunger in an oil-filled cylinder. Following an abrupt, extreme displacement, the system would come to rest in two or three cycles. Downward movements of the platform were amplified approximately 16 times, on the average, by the muscle lever which measured 4.53 in. from pivot point to writing tip. The maximum angular deviation of the lever permissible with the 5-in. polygraph was about 31°, which yielded a deflection on the record of 2.36 in. No attempt was made in reading the records to correct for the arc distortion of the stylus or for the nonlinearity introduced at the point where the recording thread was coupled to the stylus. Dynamic tests of over-all reactions of the system to suddenly imposed forces were made by dropping a 50-gm. bag of steel shot from various distances upon the top of the confinement box. On these tests, the platform was loaded with 310 gm. (the mean weight of all rats used in the experiment). These measurements revealed that the relation between the deflection of the recording pen and the fall of the 50-gm. weight was roughly linear within a height-of-fall range of 15 to 200 mm., corresponding to a pen-deflection range of 17 to 50 mm. Above and below these limits, the departures from linearity were rather marked.

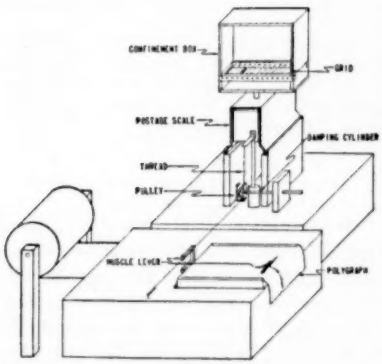

Fig. 1. Schematic drawing of apparatus for measuring magnitude of startle responses in rats to sudden loud sounds.

The presentation of the CS and UCS was controlled automatically by a group of decade-type, electronic interval-timers (6). The CS consisted of the simultaneous activation of a buzzer (Edwards, Model 1872, operated on 6 v. a.c., situated about 2.5 ft. from the confinement box) and a light (60-w. frosted lamp, 6 in. in front of the transparent side of the confinement box). The UCS for the fear reaction was a 60-cycle shock of either 70 or 80 v. (see footnote 3) applied to the grid through a series resistance of 330,000 ohms. The startle stimulus was produced by a toy pistol (Super Nu-Matic Paper Buster Gun) held about 2 ft. from the confinement box. The sound resulting from the "firing" of this gun, which reached the rat primarily through a number of holes drilled in the top of the box, was sufficiently loud and sharp to produce a marked startle reaction, yet not so loud as to require the use of a sound-shielded room. Although the gun was fired manually rather than automatically, the timing error thus introduced was quite small (estimated maximum deviation $\pm .25$ sec.). The *E* was well practiced and timed his reactions by means of a clock hand making one complete revolution in 10 sec. The exact time at which the startle stimulus occurred was registered on the

record by an electromagnetic signal marker. This marker was connected to the output of an amplifier which was activated in turn by a microphone placed near the toy pistol.

Procedure and Subjects

In broad outline, the experimental procedure described below was designed to develop a conditioned fear reaction in an experimental group of animals and to make it possible to determine whether the presence and magnitude of this anxiety state could be inferred from changes in the strength of a startle reaction *which was itself never conditioned during the experiment.*

Initial tests.—Since there appeared to be some possibility that magnitude of startle might be significantly related to body weight, each *S* was given an initial series of five presentations of the startle stimulus alone on the day preceding the actual emotional conditioning. The product-moment correlation between mean amplitude of startle on these initial tests and body weight was found to be .20 ($N = 30$) for a weight range of 232 to 368 gm. It was concluded, therefore, that body weight, as a significant variable, could be largely ignored. The variability of the unconditioned reactions to the sound was rather large, however, and it was decided to match the members of the experimental and control groups on the basis of these initial tests of reaction amplitude.

Conditioning and testing procedure.—On each of the four days following the initial test day *Ss* of the experimental group were given a series of 10 trials, seven of which were conditioning trials. On all conditioning trials, the light-buzzer CS was presented for 5 sec., with the UCS (shock) being presented for the last 2 sec. of this period.[3] This procedure was based on the assumption that the pairing of the light-buzzer combination and shock (even though no overt response was systematically reinforced) would lead to the arousal of an anticipatory fear reaction which would reach its greatest intensity at about the time the shock was normally applied. The other three trials (trials 4, 7, and 10) were tests on which the CS was presented as usual but with the startle stimulus introduced 3 sec. after the onset of the CS *in place of the shock.* The heights of all jumps to these startle stimuli were recorded. These test trials were, of course, the critical trials introduced to determine the possible facilitative effect of fear upon startle. Successive trials on any one day were administered at intervals ranging from 20 to 40 sec., the average intertrial interval being about 30 sec.

The control *Ss* were given the same number of trials, at the same rate, as were the experimental *Ss*. For the controls, however, the CS and UCS were never presented simultaneously. On alternate trials, the UCS (duration, 2 sec.) either preceded the CS by 3 sec. or followed the CS by 3 sec. By this procedure, the control *Ss* received exactly the same number of presentations of the light-buzzer and the same number of shocks as did the experimentals. This served to control for possible changes in magnitude of startle attributable to pseudo-conditioning. The controls, like the experimentals, were given three startle tests on the fourth, seventh and tenth presentations of the CS for any one day. The gun was fired 3 sec. after the onset of the light-buzzer stimulus in order to duplicate exactly the tests given the experimental animals.

Extinction trials.—On the day following the completion of the conditioning trials, all *Ss* were first given two paired (or unpaired, in the case of the controls) presentations of light-buzzer and shock, followed immediately by 12 nonreinforced presentations of the light and buzzer. On each of these extinction trials, the startle stimulus was presented at the usual time, i.e., 3 sec. after the onset of light and buzzer. Startle responses were recorded on each trial. The mean intertrial interval was again about 30 sec. Two additional series of extinction trials were given on the next two days in exactly the same manner, with the single exception that no more shocks were given the members of either group.

Subjects.—Thirty hooded, male rats, divided into two equal groups, were used. Their ages ranged from 90 to 120 days, and all had been previously used in a bar-pressing experiment which did not involve shock.

Results

Acquisition.—The results obtained from measurements of startle-response amplitude on each of the four conditioning days and on the preceding "matching" day are presented in Table I. The *S's* scores for the four conditioning days were means of the startle responses elicited on the three test trials of each day. The values

[3] The strength of shock given each *S* was either 70 or 80 v., depending upon its degree of sensitivity. Rough determinations of the reactivity level of all *Ss* were made on the initial test day following the five startle trials. The assignment of shock voltages was based upon visual inspections of the average amplitude of activity tracings provided by the recording system when shock was applied. The number of *Ss* receiving the 80-v. shock was the same in both groups.

TABLE I

MAGNITUDES OF STARTLE RESPONSES ON A PRETEST DAY USED FOR MATCHING GROUPS AND ON FOUR SUCCEEDING DAYS DURING WHICH EXPERIMENTAL *Ss* WERE CONDITIONED

Group	Measure in mm.	Matching Pretest	Successive Conditioning Days			
			1	2	3	4
Experimental	*M*	12.3	20.7	23.9	25.5	25.1
	Mdn	9.6	19.7	23.3	24.3	25.0
	SD	9.9	14.8	14.0	14.0	13.5
Control	*M*	11.8	13.7	15.6	15.0	11.9
	Mdn	9.2	11.7	8.3	11.7	10.3
	SD	10.4	9.9	13.2	13.7	10.0

listed in the "Matching Pretest" column were computed for the experimental and control groups after matching on the basis of the initial day's tests. These values were included in the table to indicate the approximate magnitude of the startle response prior to the conditioning of the emotional reaction.[4]

From an inspection of the tabulated values and of the graphic plot of the medians in Fig. 2, it will be seen that the experimental *Ss* showed a marked increase in magnitude of startle response as a result of the anxiety-conditioning trials. By contrast, the responses of the control *Ss* appeared to show no such systematic change in magnitude, the mean for the last day of training being almost identical with that obtained on the initial test day. The *net* over-all increase in mean response magnitude between the pretest day (Day 0) and Day 4 shown by the experimental group as compared with the controls is highly significant ($t = 3.3$, $df = 14$, $p = <.01$). The tentative conclusion may be drawn, therefore, that the paired presentation of the CS and the shock led to the development of a conditioned fear state that functioned to facilitate the *Ss*' startle reactions to the sound.

In addition to the gross finding that magnitude of startle was intensified by fear, it is also apparent that the process of intensification was, within lim-

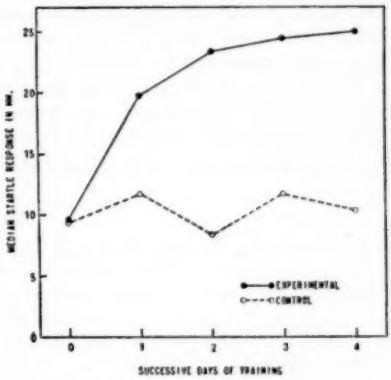

FIG. 2. Curves showing the differential facilitative effect of fear conditioning upon magnitude of startle response to an explosive sound. The training procedure was designed to induce a state of fear in the experimental animals but not in the controls. The values plotted at the zero point on the abscissa were obtained from measurements of startle on the day prior to the initiation of training.

[4] This measure of "zero conditioning" is not, it is recognized, entirely adequate since it was obtained from reactions to sound alone. A more appropriate measure would have been magnitude of startle response to sound *plus* light and buzzer. Thus it is possible that the presence of the light and buzzer might have some inherent dynamogenic effect upon startle and that our zero point is therefore spuriously low. We would reject this possibility, however, on the grounds that the control group's scores for successive days (especially the medians) do not deviate markedly from those of the matching day in spite of the additional facilitative effect that could have been produced by the residual emotional effects of shock at the time the sound was given. In any event, it is obvious that the experimental and control groups were highly similar in terms of their startle responses prior to training, and it is doubtful that the addition of the light and buzzer would have had a *differential* effect upon their reactions.

its, a progressive one. The curve of median values, plotted in Fig. 2, resembles rather closely more conventional negatively accelerated curves of conditioning. It is also evident, however, that the growth of fear is an extremely rapid process when compared with the speed of acquisition of a skeletal response in a classical conditioning situation. From our data, it would appear that the reaction was quite strongly conditioned by the end of the first day. This observation regarding the rapidity of emotional conditioning is further supported by the finding that on the initial startle trial of Day 1, given after only three paired presentations of light-buzzer and shock, the experimental group showed a mean startle response of 21.7 mm., whereas the comparable mean for the controls was only 8.5 mm. It is conceivable, therefore, that if a large number of *Ss* were run so as to stabilize plots of reactions on individual test trials, the curve reflecting the acquisition of fear in this situation would exhibit an even more precipitous rise than that suggested by Fig. 2.

Qualitative observations of the behavior of the *Ss* during the conditioning series revealed that their reactions to the CS resembled somewhat those reported by Upton (**16**) for guinea pigs and by Wever (**17**) for cats. In the case of the experimental rats, it was noted that quite early in the conditioning series the onset of the CS was often followed by a marked reduction in general activity level accompanied by rapid, shallow breathing reminiscent of the "flutter phenomenon." The controls, on the other hand, were more prone to exhibit this behavior during the interval *following* the cessation of the CS. For them, it appeared as if the termination rather than the onset of the CS had become the critical environmental change arousing fear. This observation is supported in part by additional data included in the Discussion.

Extinction.—On the assumption that the data obtained during the acquisition trials reflected the growth of a conditioned fear reaction, it appeared reasonable to expect that with the omission of shock, the reaction would show such additional phenomena of classical conditioning as extinction and spontaneous recovery. In order to examine this possibility, three days of extinction trials were introduced following the acquisition series. The descriptive statistics obtained from the extinction-trial data are presented

TABLE II

MAGNITUDES OF STARTLE RESPONSES FOR SUCCESSIVE BLOCKS OF THREE TRIALS DURING EACH OF THREE DAYS OF EXTINCTION TRIALS

Group	Measure	Successive Blocks of Three Extinction Trials Each											
		Day 1				Day 2				Day 3			
		1	2	3	4	5	6	7	8	9	10	11	12
Experimental	*M*	25.0	18.4	11.5	12.4	17.7	15.0	12.3	11.6	13.0	16.1	11.7	11.8
	Mdn	19.3	16.7	10.3	7.0	16.3	9.0	8.0	5.0	7.7	11.3	7.3	11.3
	SD	16.9	11.9	8.9	10.1	14.0	13.8	13.0	13.1	10.6	16.2	11.3	12.2
Control	*M*	12.7	11.7	7.9	10.7	8.6	5.6	7.0	6.6	9.2	7.3	7.3	6.5
	Mdn	11.7	8.0	8.7	7.0	7.3	3.3	6.7	6.3	5.0	8.3	6.3	4.3
	SD	10.2	10.0	6.0	9.3	5.4	5.6	4.6	6.1	8.8	5.3	5.7	5.6

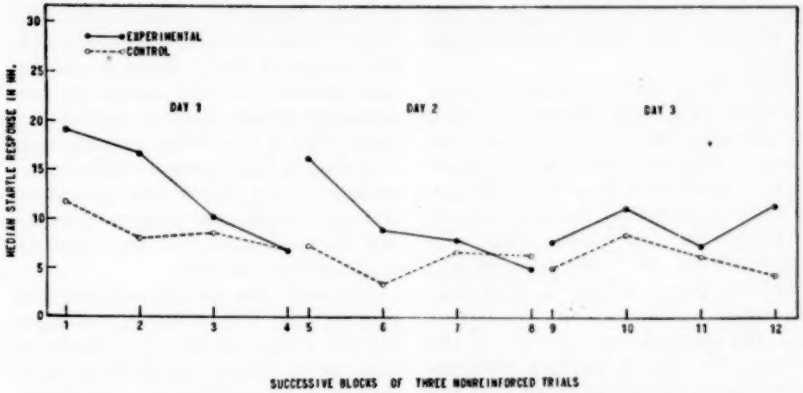

FIG. 3. Magnitude of startle responses exhibited by the experimental (fearful) and control (nonfearful) groups on three successive days during which the previously conditioned, fear-arousing stimulus was repeatedly presented without being followed by the unconditioned stimulus of electric shock

in Table II. These values were computed from distributions of individual scores, each of which was a mean of three successive extinction trials. A plot of the median values from the table is provided in Fig. 3 to facilitate visual inspection.

It will be seen from the first day's data that the vigor of the startle reaction exhibited by the experimental *Ss* underwent a progressive decline as a function of the number of nonreinforced trials. The control *Ss* also showed a slight decrement in response strength which could have been due to adaptation, or to the extinction of a small amount of anxiety, or to both. Negative adaptation could also, of course, have been a factor contributing to the fall in the experimental-group curve. The difference between the means for the experimental group on the first and fourth blocks of trials was highly significant ($t = 3.13$, $df = 14$, $p = <.01$). Since the controls also showed a slight drop, the *net* decrease for the experimental group relative to the controls was of course reduced, and was found to be significant at about the $.10 > p > .05$ level of confidence. It seems reasonable to infer from this evidence, then, that the fear which was presumably responsible for the intensification of the startle response, declined in strength as the result of the operation of omitting the shock.

From the experimental-group data of the second extinction day it is evident that the 24-hr. interval separating the fourth and fifth blocks of trials led to the spontaneous recovery of the startle response and hence, by inference, to the recovery of the extinguished fear response. The means for the control *Ss*, by contrast, showed a decrement rather than an increase, with the medians showing no appreciable change. A statistical evaluation of the net mean change from Blocks 4 to 5 for the experimental versus the control *Ss* yielded a t of 2.00, with a p between .05 and .06. This suggestive evidence for the differential effects of the passage of time upon the responses of the two groups is in accord with the notion that fear, as reflected by startle responses, is a conditionable state or

response having functional properties in common with other conditionable reactions.

With further nonreinforced trials on the second day, the indicator reaction for the experimental group again declined in much the same manner as on Day. 1. Spontaneous recovery, though in reduced amount as would be expected from other conditioning data, was again shown by the experimental group at the start of Day 3. This finding tends to support the conclusion that the spontaneous recovery of the second day was a genuine phenomenon, but it should be noted that the means (though not the medians) for the control group were also greater at the start of Day 3 than at the end of Day 2. No clear indication of the occurrence of further progressive extinction was provided by the third day's data.

Discussion

The results of the present experiment have provided additional support for the supposition that the paired presentation of a neutral and a noxious stimulus results in the acquisition, by the neutral stimulus, of the capacity to elicit an emotional response of fear or anxiety.[5] That this conditioned fear state appears to possess motivating or energizing properties akin to those of primary drives is suggested by the finding that at the time fear was presumably aroused, the innate startle responses of rats to a loud sound were considerably intensified. Moreover, since the degree of intensification of startle varied concomitantly with operations known to affect strength of conditioning, and since the startle reaction was itself never directly conditioned, it may be assumed that the effect of these operations was to change the degree of fear. Thus it can be inferred that fear was increasing progressively with number of conditioning trials, that it was being extinguished, and that it had spontaneously recovered, from the observation that these classical conditioned-response properties were exhibited by the nonconditioned startle reaction.

Although the startle response has been termed *nonconditioned* throughout this study, the fact that this reaction *can* be conditioned (**7, 8**) makes it necessary to consider the possibility that the obtained results could be attributed to learned modifications of the startle tendency instead of to the energizing property of fear. In the present experiment, the jump response was elicited on 12 occasions (three times on each of four training days) in the presence of the CS, and it would be on these trials, therefore, that direct conditioning of startle might have taken place. Thus it might be supposed that the response was conditioned through the action of sheer contiguity of stimulus and response or because of the reinforcement provided by the offset of the sound stimulus. Although this is a logical possibility, it does not provide an acceptable explanation of the differences between the experimental and control groups. Inasmuch as both groups were given precisely the same number and kind of test trials, they had exactly the same opportunity for the startle to become conditioned in the suggested manner.

An alternative and more plausible hypothesis is that the tendency to jump to the shot in the presence of the light and buzzer was increased because of the *reinforcing effect of fear reduction* attending the cessation of the CS, rather than the energizing effect of the anxiety elicited by the CS. The fact

[5] This statement involves no implicit hypotheses as to the basic conditions for this acquisition. It in no way implies, for example, that the mere contiguity of the two stimuli is regarded as sufficient for the learning of the emotional reaction.

that the responses of the controls showed no significant increase is perfectly compatible with this proposal since, for them, the CS should have led to little or no anxiety as compared with the experimentals. Although this hypothesis differs from the major one of the present paper, it leads to precisely the same conclusions with respect to the growth and decay of fear.

Evidence can be cited, however, which favors the conclusion that the major determinant of the increase in startle was the dynamogenic rather than the reinforcing action of fear. Thus it may be recalled that the average jump of the experimental *Ss* to the *first* presentation of the startle stimulus, introduced after three conditioning trials, was 21.7 mm., almost twice as large as the mean response of the same *Ss* on the matching trials of the previous day. Since the startle stimulus had not, prior to this first test, been followed by fear reduction, it appears unlikely that this marked increase in strength could have been due to a learned modification of the startle tendency. We conclude, therefore, that although the possible role of fear reduction cannot, and should not, be completely dismissed, a more defensible hypothesis is that the observed effects were primarily due to the energizing function of fear.

A still further interpretation of the present findings might rest upon the possibility that the experimental *Ss*, but not the controls, were learning to adopt a specific posture in response to the CS. If this were the case, it is conceivable that the enhancement of the startle to the shot could be explained in terms of the sheer "mechanical advantage" provided by the learned postural adjustment. Successive changes in the posture, and hence in its capacity to facilitate jumping, would be presumed to have taken place during the training and extinction trials. Since no precise records were made of the animals' postural or gross muscular adjustments to the light and buzzer, it is possible that such adjustments could have occurred.

With respect to this hypothesis, it may be noted initially that fear has been defined simply as a conditioned anticipatory response to impending painful stimulation. This response, or response pattern, may be assumed to produce intense stimulation, and thereby, a state or process having certain of the properties of primary drives. In this formulation no attempt has been made to specify whether the anticipatory responses need be central or peripheral, skeletal or visceral. Nor have any assumptions been made as to whether the physical mechanisms underlying the intensifying function of fear are hormonal, neural, or mechanical. The basic hypothesis that fear acts as an energizer would not, therefore, be invalidated by the observation that postural adjustments were present or by the finding that these adjustments were facilitative in a purely mechanical fashion.

It should be observed, however, that perhaps not *all* responses elicited by the cues associated with a noxious stimulus should be considered as fear reactions. Some of these anticipatory responses may produce little stimulation and therefore have little or no drive value, and it is conceivable that some conditioned postural adjustments are of this sort. It is important to note that if the intensification of startle is due solely to the mechanical advantage inhering in such postures then the range of responses susceptible to enhancement as a result of this reaction is probably severely restricted. Response patterns having motivational properties are usually assumed to have the capacity of intensifying a rather

wide variety of reactions. It is doubtful that there would be much utility in ascribing drive characteristics to a response if its intensifying action were limited to a single other reaction. Whether the anticipatory responses in the present situation have such restricted effects is, of course, a question that must be answered by further investigation.

It may be recalled that one of the rather casual observations of the present study was that the going-off of the CS rather than its onset appeared to be the critical fear-arousing event for the *Ss* of the control group. Quantitative data supporting this supposition were obtained from eight *Ss* following the completion of the main experiment. In this exploratory study, four rats from each of the original experimental and control groups were given ten reconditioning trials on each of two successive days. The procedure was identical with that of the principal study except for the startle test trials. On these trials, the startle stimulus was presented 3 sec. *after the termination* of the CS instead of 3 sec. after its onset, as had previously been the case. On the assumption that fear should have been greatest at about the time the shock was normally received, it follows that 3 sec. after the termination of the CS, the controls, that were shocked on alternate trials at that time, should have had a higher degree of fear than the experimentals. That this was apparently the case is indicated by the fact that the mean response of the controls on the six test trials of the two days was 13.9 mm., as compared to a mean of 6.9 mm. for the experimental *Ss*.

Although this reversal in the relative response strengths of the two groups is reasonably striking, the finding must be regarded as quite tentative. It may be remarked, incidentally, that the ease with which it was possible to make this additional observation would appear to constitute one of the major advantages of the present method. To have attempted to demonstrate the fear-arousing properties of the cessation of a stimulus by methods involving reinforcement through fear reduction would have been far more laborious.

Finally, it should be pointed out that the present experiment is to a certain degree similar to the "sensitization" study of Prosser and Hunter (**15**). They observed that a weak auditory stimulus originally incapable of eliciting a startle response of the gastrocnemious muscle of the rat would acquire the capacity to do so after it had been paired with shock. Their conclusion that excitability is increased by shock and therefore facilitates startle is consistent with our findings that startle is intensified by fear. In their experiment, however, no obvious additional stimulus capable of arousing a conditioned fear state preceded the weak sound. Hence, it seems likely that their observed sensitization was due to the effects of *residual* excitement from preceding shocks rather than to a *conditioned* excitement as in the present experiment.

Summary and Conclusions

1. Previous experimental demonstrations of the motivational aspects of fear or anxiety as a determinant of behavior have been based primarily upon the reinforcing property of fear reduction. Although this approach has been quite successful, a number of practical disadvantages can be discerned in the method. The present paper describes an alternative method of studying fear, which rests upon assumptions concerning the energizing

rather than the reinforcing properties of fear.

2. Rats were placed in a stabilimeter-like device and given a series of training trials, each of which consisted in the presentation of a neutral stimulus and a shock. For an experimental group, the two stimuli were paired in a manner designed to produce a conditioned pain response (fear or anxiety) to the neutral stimulus; for a control group, the temporal intervals between the stimuli were such as to reduce or prevent the formation of fear. Under neither procedure was any overt response systematically reinforced. Following the conditioning series, extinction trials (CS presented alone) were given on each of three successive days. To test for the presence of fear, a loud, sharp sound was substituted at the time shock was customarily applied (to the experimental *Ss*) and the magnitude of the startle response was recorded. On the assumption that the startle reaction would be a function of the strength of the conditioned fear state, it was predicted that vigor of startle would vary concomitantly with operations designed to produce acquisition, extinction, and spontaneous recovery of fear.

3. The results for the training series showed a significant increase in the average startle response of the experimental *Ss* relative to the controls (whose reactions changed but little during these trials). Data from the series of extinction trials revealed that the startle response in the experimental group declined progressively during the nonreinforced trials of the first day, recovered spontaneously during the interval separating the first and second days, declined again on the second day, and tended to show recovery again between the second and third days.

4. It is concluded that the results of the experiment provide further support for the assumption that the pairing of a neutral and a noxious stimulus leads to the development of a conditioned fear reaction having motivational concomitants. It is further held that the momentary strength of this reaction can be inferred from the extent to which some other reaction, in itself never directly conditioned to the fear cues, is intensified at the time of fear arousal.

(Manuscript received May 12, 1950)

References

1. Brown, J. S. A note on a temporal gradient of reinforcement. *J. exp. Psychol.*, 1939, **25**, 211–227.
2. Brown, J. S., and Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
3. Estes, W. K., and Skinner, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, **29**, 390–400.
4. Farber, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
5. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
6. Hunter, T. A., and Brown, J. S. A decade-type electronic interval-timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
7. Hunter, W. S. Muscle potentials and conditioning in the rat. *J. exp. Psychol.*, 1937, **21**, 611–624.
8. Landis, C., and Hunt, W. A. *The startle pattern.* New York: Farrar and Rinehart, 1939.
9. May, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66–77.
10. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.
11. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
12. Miller, N. E., and Dollard, J. *Social learning and imitation.* New Haven, Conn.: Yale Univ. Press, 1941.

13. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
14. Mowrer, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, **27**, 497–516.
15. Prosser, C. L., and Hunter, W. S. The extinction of startle responses and spinal reflexes in the white rat. *Amer. J. Physiol.*, 1936, **117**, 609–618.
16. Upton, M. The auditory sensitivity of guinea pigs. *Amer. J. Psychol.*, 1929, **41**, 412–421.
17. Wever, E. G. The upper limit of hearing in the cat. *J. comp. Psychol.*, 1930, **10**, 221–233.

THE INTELLIGIBILITY OF SPEECH AS A FUNCTION OF THE CONTEXT OF THE TEST MATERIALS [1]

BY GEORGE A. MILLER, GEORGE A. HEISE, AND WILLIAM LICHTEN

Harvard University

For many years communication engineers have used a psychophysical method called the "articulation test" (**2, 3**). An announcer reads lists of syllables, words, or sentences to a group of listeners who report what they hear. The articulation score is the percentage of discrete test units reported correctly by the listeners. This method gives a quantitative evaluation of the performance of a speech communication system.

There are three classes of variables involved in an articulation test: the *personnel*, talkers and listeners; the *test materials*, syllables, words, sentences, or continuous discourse; and the communication *equipment*, rooms, microphones, amplifiers, radios, earphones, etc. The present paper is directed toward the second of these three classes of variables, the test materials. The central concern can be stated as follows: Why is a stimulus configuration, a word, heard correctly in one context and incorrectly in another?

Three kinds of contexts are explored: (*a*) context supplied by the knowledge that the test item is one of a small vocabulary of items, (*b*) context supplied by the items that precede or follow a given item in a word or sentence, and (*c*) context supplied by the knowledge that the item is a repetition of the immediately preceding item. All three kinds of context enable the listener to limit the range of alternatives from which he selects his response. A word selected from a small vocabulary must be one of the few words agreed upon in advance. A word in a sentence must be one of the relatively few words that make a reasonable continuation of the sentence according to grammatical rules agreed upon in advance. A repeated word must be one of the few words similar to the word just heard. Not anything can happen, and the listener can set himself to make the required discrimination. Context, in the sense the word assumes here, is the *S's* knowledge of the conditions of stimulation. The experimental problem is to vary the nature and amount of this contextual knowledge in order to study its influence upon perceptual accuracy.

Equipment and Procedure

The apparatus consisted of components from military communication equipment used during the recent war. The output voltage of a carbon microphone was amplified and delivered to the listener's dynamic earphones. The talker monitored his speaking level with a volume indicator (VU meter) that responded to the voltage generated at the output of the amplifier. A random noise voltage, with a spectrum that was relatively uniform from 100 to 7,000 cps, was introduced at the listener's earphones. The signal-to-noise ratio (S/N) was varied by holding the average voice level constant and changing the level of the noise. The S/N was measured by a vacuum tube voltmeter across the terminals of the earphones, and the measurements reported in

[1] This research was carried out under Contract N5ori-76 between Harvard University and the Office of Naval Research, U.S. Navy (Project NR147-201, Report PNR-74). Reproduction for any purpose of the U.S. Government is permitted.

the following pages represent the ratio in decibels of the average peak deflection of the meter for the words (in the absence of noise) to the level of the noise in the 7,000-cycle band. A S/N of zero db means, therefore, that the electrical measurements indicated the two voltages, speech and noise, were equal in magnitude. Since the earphones transduce frequencies only up to about 4,500 cps, however, the acoustic level of the noise was about 2 db lower than these electrical measurements indicate. The over-all acoustic level of the voice at the listener's ears was approximately 90 db re .0002 dyne/cm[2].

The speech channel was not a high quality system. Only the speech frequencies between 200 and 3,000 cps were passed along to the listener.[2]

Only two *Ss* were used throughout the experiments. Both had normal auditory acuity, and both were familiar with the design and theory of the experiments. The *Ss* were located in different rooms, connected only by the communication channel described above, and they alternated as talker and listener. Some particular S/N was set up in the channel, and the talker proceeded to read a list of test items. These items were pronounced after a carrier sentence, "You will write. . . ." During this carrier sentence the talker adjusted his voice level to give the proper deflection of the monitoring VU meter, and then the test item was delivered with the same degree of effort. This procedure preserves the inherent variability of English words—the word "peep" has much less acoustic energy than the word "raw" when both words are pronounced with equal emphasis by a normal talker. By monitoring the carrier sentence rather than the test item, the relative intensities of the speech sounds are preserved in a natural fashion. The listener then recorded the item on a test blank, and these test sheets were later graded and the scores converted to percentages.

Importance of Test Materials

The kind of speech materials used to test communication systems is an important variable determining the results of the tests. Figure 1 illustrates how much difference the test materials can make. These three functions were obtained for the communication channel and the personnel described above. The test materials used for these three functions were the following.

(*a*) The *digits* were pronounced *zero, wun, too, thuh-ree, four, f-i-i-v, six, seven, ate, niner.* (*b*) The *sentences* were those constructed at the Psycho-Acoustic Laboratory (1). A sentence consists of five major words connected by auxiliary "of's," "the's," etc. The score shown in Fig. 2 represents the percentage of these major words heard correctly. (*c*) The *nonsense syllables* used were also those published by Egan (1). To standardize the pronounciation and recording of the nonsense syllables, an abbreviated phonetic symbolism was used.

The values of S/N necessary for 50 per cent correct responses are approximately -14 db for digits, -4 db for the individual words in a sentence, and $+3$ db for nonsense syllables. At a S/N where practically no nonsense syllables were recorded correctly, nearly all the digits were correctly communicated. Differences of this magnitude require explanation. What differences among these spoken stimuli make some easy to hear and others quite difficult? A list of perceptual aspects—rhythm, accent, grouping, meaning, or phonetic composition—can be suggested.

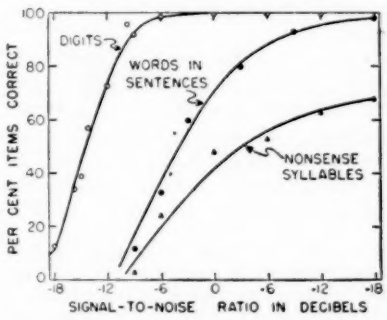

FIG. 1. Relative intelligibility of different test materials

[2] For the convenience of those who may wish to apply one of the several schemes for predicting articulation scores, the frequency response of the system may be obtained by ordering Document 3250 from American Documentation Institute, 1719 N Street, N. W., Washington 6, D.C., remitting $1.00 for microfilm (images 1 in. high on standard 35 mm. motion picture film) or $1.00 for photocopies (6 × 8 in.) readable without optical aid.

Our experiments indicate, however, that these various characteristics of the stimulus that *did* occur are less important than the characteristics of the stimuli that *could* have occurred but didn't. The most important variable producing the differences is the range of possible alternatives from which a test item is selected. A listener's expectation (or, more precisely, his freedom of choice) is determined by the context in which the particular phonetic pattern occurs. When digits are used, the listener can respond correctly with a marginal impression of the relatively intense vowel sound alone, because all the digits, with the exception of *five* and *nine*, have different vowels. Since the alternatives are thus limited, the digit is interpreted correctly, although the same acoustic stimulus is quite ambiguous when the alternatives are not so limited. With nonsense syllables, however, this limitation of possibilities is far less helpful; the listener feels that anything can happen. To record the nonsense syllable correctly, a listener must perceive each phoneme correctly, and the perception of one phoneme in a syllable does not give a clue to the other phonemes in the same syllable. Not only must the listener hear the vowels correctly, but the less intense consonant sounds must also be distinguished.

Size of Test Vocabulary

An articulation test is a rather unusual combination of the familiar psychophysical procedures. The experiment requires the listener to select, not one out of two or three, but one out of several thousand alternative responses. Thus the number of alternatives involved becomes an interesting variable.

Suppose we try to adapt spoken stimuli as closely as possible to the traditional method of constant stimuli. To this end we might use a single speech sound or a single syllable as the stimulus, present this speech unit at various intensities, and ask *S* to report whether or not he heard each presentation. This procedure determines a threshold of audibility for the particular speech unit. The practical value of this isolated datum is negligible. The experiment must be repeated for all the forty or fifty different speech sounds or the thousands of different syllables of English. And then we know only about audibility, not intelligibility.

Consider this distinction between audible speech and intelligible speech. It is intuitively clear that the words *audible* and *intelligible* are not synonyms, and listeners give reliably higher thresholds when asked to make continuous discourse "just understandable" instead of "just audible" (**2**). The crux of the difference is that intelligibility involves a complex discrimination and identification, whereas audibility is simply a discrimination of presence or absence.

It seems reasonable, therefore, to call a speech unit intelligible when *it is possible for an average listener with normal hearing to distinguish it from a set of alternative units*. By a speech unit is meant any combination of vocal noises—phonemes, syllables, words, phrases, sentences. The act of distinguishing can take various forms—repeating the unit, writing it down, pointing to it, behaving in accordance with its content, etc. The critical part of this definition concerns the set of alternative speech units from which the particular unit is selected. This part of the definition reduces intelligibility to discriminability, and avoids the questions of semantic rules and meaning. Discriminability is a function of the number of alternatives and

the similarities among them. The word "loot" is easily discriminated if all the alternatives are trisyllabic, but difficult to distinguish, other things being equal, in a set of alternatives that includes "boot," "loop," "jute," "lewd," "mute," "loose," etc.

An articulation test is analogous to a test of visual acuity where the percentage of correct judgments of a fixed set of test figures is plotted as a function of the level of illumination. A differential judgment is required under various favorable and unfavorable conditions. In such an experiment we determine the most unfavorable conditions under which the discrimination can be made, rather than the most unfavorable conditions under which the presence of the stimuli can be detected. These are clearly different thresholds and correspond to what we have called the thresholds of intelligibility and of audibility.

A difficult discrimination quickly becomes impossible as the conditions are made unfavorable, whereas an easy and obvious difference remains noticeable almost as long as the stimuli can be detected. The discrimination of a difference of 3 cycles in frequency, for example, is fairly accurate under favorable conditions—at 1,000 cps and 100 db. If the intensity is progressively reduced, however, such a small difference becomes imperceptible. For a simpler discrimination, say 30 cycles difference in frequency, the listener can respond accurately at all intensities down to 5 or 10 db above the threshold of audibility.

The situation is manageable so long as we have some index of the difficulty of the discrimination. Thus, in the tonal example, the difficulty can be gauged by the size of the difference in frequency. With the articulation test, however, such an index is not available. We could utilize known differences in the spectra of the sounds to construct an index of the distance between speech sounds, but this index is not yet available. For the present we must approach the problem in a simpler way.

Imagine a many-dimensional space with a separate coordinate for each one of the different frequencies involved in human speech sounds. Along each coordinate plot the relative amplitude of the component at that frequency. In this hyperspace each unique speech sound is represented by a single point. Each point in the hyperspace represents a single acoustic spectrum. The group of similar sounds comprising a phoneme is represented by a cluster of points in the hyperspace. If a language utilized only two different phonemes, the hyperspace could be split into two parts, one for each phoneme. The distance between the two phonemes could be made as large as the vocal mechanism permits, and discrimination would be relatively easy. But suppose the number of different phonemes in the language is increased from two to ten. With ten different phonemes the hyperspace must be divided into at least ten subspaces, and the average distance between phonemes must be smaller with ten phonemes than it is with two. The discriminations involved must be correspondingly more precise. If the number of alternative phonemes is increased to a thousand, then the listener is required to make even more precise discriminations.

In other words, the ease with which a discrimination of speech sounds can be made is limited according to the number of different speech sounds that must be discriminated. From this line of reasoning it follows that the number of alternatives can be used to gauge the difficulty of discrimination. This argument has been developed by Shannon (5) to give a measure of the amount of information in a message. The interesting aspect of this index of difficulty, or of amount of information, is that it does not depend upon the characteristics of the particular item,

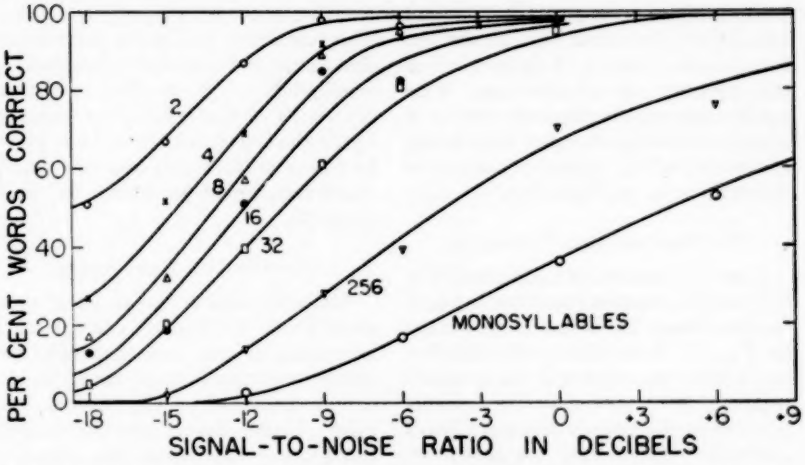

FIG. 2. Intelligibility of monosyllables as a function of the size of the text vocabulary. (Data are not corrected for effects of chance.)

but upon the range of items that *could* occur.

The range of alternatives was used as the experimental variable in the following way. The listener was informed that each test word would be one of the items from a given restricted vocabulary. The size of the test vocabulary was alternatively 2, 4, 8, 16, 32, or 256 words. The talker always spoke one of the words from the prearranged list.

The words used in the restricted vocabularies were chosen at random from the list of phonetically balanced monosyllables published by Egan (2). For the two-alternative vocabulary, different pairs of words were chosen and typed on the listener's answer sheet. The talker read one of the pair, and the listener checked the item he heard. A similar procedure was used for the four- and eight-word vocabularies. For the 16-, 32-, and 256-word vocabularies the listeners had a list of all the words before them, and studied this list until they made their choice. The choice was recorded and a signal given to the talker to proceed to the next item. The *S*s studied carefully the particular list used in any test and arranged the words according to the vowel sounds before the tests began.

The results are summarized in Table I and in Fig. 2. Included with the data for restricted vocabularies are data for words from the original list of 1,000 monosyllables, obtained with no list of choices available to the listener. When these data are corrected for chance, the two-word vocabulary gives a threshold (50 per cent of the words correct) at −14 db, the 256-word vocabulary gives a threshold at −4 db, and the unrestricted list of

TABLE I

PER CENT WORDS CORRECT IN ARTICULATION TESTS WITH VOCABULARIES OF VARIOUS SIZES

S/N in db	Size of Vocabulary							
	2	4	8	Digits	16	32	256	Monosyllables
−21	49							
−18	51	27	17	13	13	5		
−15	67	52	32	38	19	20	2	
−12	87	69	57	73	51	39	14	3
− 9	98	92	89	92	85	61	28	
− 6		94	95	99	82	81	39	17
− 3		96						
0				100	97	95	70	37
+ 6				100			76	53
+12				100			90	70
+18								82

monosyllables gives a threshold at +4 db. With the same test words the threshold is changed 18 db by varying the number of alternatives. This result supports the argument that it is not so much the particular item as the context in which the item occurs that determines its intelligibility.

Context of the Sentence

A word is harder to understand if it is heard in isolation than if it is heard in a sentence. This fact is illustrated by Fig. 3. Sentences containing five key words were read, and the listener's responses were scored as the percentage of these key words that were heard correctly. These data are shown by the filled circles in Figs. 1 and 3. For comparison, the key words were extracted from the sentences, scrambled, and read in isolation. The scores obtained under these conditions are shown by the open circles of Fig. 3. The removal of sentential context changes the threshold 6 db.

The effect of the sentence is comparable to the effect of a restricted vocabulary, although the degree of restriction is harder to estimate. When the talker begins a sentence, "Apples grow on ———," the range of possible continuations is sharply restricted. This restriction makes the discrimination easier and lowers the threshold of intelligibility. A detailed statistical discussion of the restrictions imposed by English sentence structure is given by Shannon (**5**), and is used in a simple recall experiment by Miller and Selfridge (**4**).

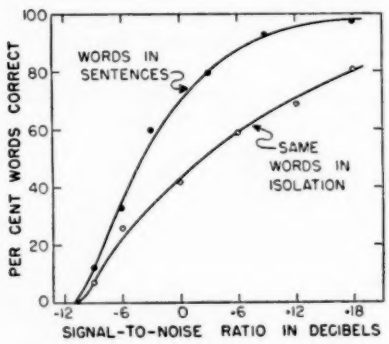

Fig. 3. Effect of sentence context on the intelligibility of words

Effects of Repetition

When an error occurs in vocal communication, the listener can ask for a repetition of the message. The repeated message is then heard in the context provided by the original message. If the original message enabled the listener to narrow the range of alternatives, his perception of the repeated message should be more accurate. A series of tests were run with various kinds of test materials to evaluate the importance of the context of repetition. These tests were run with automatic repetition of every item and, also, with repetitions only when the listener thought he had not received the test item correctly.

The improvement in the articulation scores obtained with automatic and with requested repetitions was found to be about the same. A slight but insignificant difference was found in favor of the requested repetition, and if we add to this the savings in time achieved by omitting the unnecessary repetitions, the requested repetition is clearly superior.

The advantage gained by repetition is small for all types of test materials. In Fig. 4 data are given for the effects of repeating automatically the monosyllabic words. The difference in threshold between one presentation and three successive presentations is only 2.5 db. Similar data for words heard in sentences show a shift of 2 db, and for digits, 1.5 db.

These results indicate that the improvement that can be achieved by the simple repetition of a message is

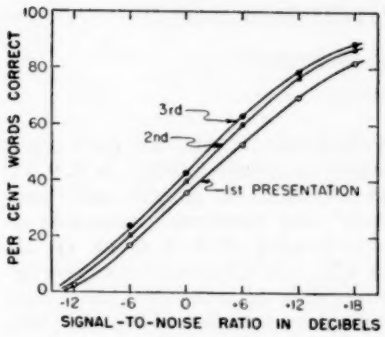

FIG. 4. Effects of repetition of test words on the articulation score

slight. The repeated message contains approximately the same information, and the same omissions, that the original message contained. If the listener thinks he heard the word correctly, he persists in his original response, whether it is right or wrong. If he thinks he heard the word incorrectly, he does not use this presumably incorrect impression to narrow the range of possibilities when the item recurs. In any case, no strong factor is at work to improve the accuracy on repeated presentations, and so we obtain only the slight improvement indicated in Fig. 4.

The results indicate that far more improvement in communication is possible by standardizing procedures and vocabulary than by merely repeating all messages one to two times.

In general, therefore, the results are in qualitative agreement with the mathematical theory of communication presented by Shannon (**5**). A precise quantitative comparison of the data with the theory cannot be made in the absence of trustworthy information about the distributions of errors. Seemingly reasonable assumptions about the error distributions give results consistent with theoretical predictions, but a more thorough study would be rewarding. For a given signal-to-noise ratio the listener receives a given amount of information per second (according to Shannon's definition), and articulation scores can be predicted for different types of test materials on the basis of the average amount of information needed to receive each type of test item correctly.

SUMMARY

Articulation tests showed the effects of limiting the number of alternative test items upon the threshold of intelligibility for speech in noise. The number of alternative test items was limited by providing three kinds of context: (*a*) restricting the size of the test vocabulary, (*b*) using the words in sentences, and (*c*) repeating the test words. Differences among test materials with respect to their intelligibility are due principally to the fact that some materials require more information than others for their correct perception. The relative amount of information necessary for a given type of item is a function of the range of alternative possibilities. As the range of alternatives increases, the amount of information necessary per item also increases, and so the noise level must be decreased to permit more accurate discrimination.

(Manuscript received April 24, 1950)

REFERENCES

1. EGAN, J. P. *Articulation testing methods, II.* OSRD Report No. 3802, February, 1942. (Available through Office of Technical Services, U.S. Department of Commerce, Washington, D.C., as PB 22848.)
2. EGAN, J. P. Articulation testing methods. *Laryngoscope*, 1948, **58**, 955–991.
3. FLETCHER, H., AND STEINBERG, J. C. Articulation testing methods. *Bell Syst. Tech. J.*, 1929, **8**, 806–854.
4. MILLER, G. A., AND SELFRIDGE, J. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
5. SHANNON, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379–423, 623–656.

SENSORY CUES IN PITCH JUDGMENT

BY LAWRENCE M. BRAMMER[1]

Stanford University

Three problems were investigated in this study: (1) Do visual, kinesthetic, and tactual cues added to hearing make pitch judgments more accurate than hearing cues alone? (2) Does the distribution of pitch judgments for violinists claiming absolute pitch differ in accuracy from judgments of those not claiming such ability? (3) What devices are used by violinists to name tones accurately?

Little is known about the importance of cues, other than hearing, which may enhance pitch-judging accuracy. Mursell (**10**) concluded, after reviewing studies measuring larynx-type movements, that the feeling of tension in the vocal muscles does not increase accuracy. Wyatt (**24**), in a study on pitch discrimination, states that:

> It seems probable that proficiency in a pitch discrimination test may be affected by many factors which are quite remote from simple auditory sensation and which may involve not just the auditory sensorium, but possibly even the entire organism (p. 54).

Cheslock (**5**) and Gough (**6**) assert that kinesthetic cues make pitch judgments more accurate.

The last two problems of this study have a long history. The definition and investigation of the absolute pitch phenomenon are discussed at length in a critical review by Neu (**11**).

Absolute pitch is defined as the ability to name tones from memory only. The error allowed varies from .01 to 1 semitone (**2, 8, 9, 15, 16, 17, 18, 19**). A review of these studies on pitch judgment suggests the following generalizations: (1) Absolute pitch is primarily an acquired ability (**6, 8, 9, 11, 13, 20, 22**). (2) Timbre and "tone body" are mentioned frequently in pitch-judging (**2, 3, 4, 9, 10, 11, 12, 13, 17**). (3) Unmusical observers improve their accuracy in a short time through systematic training (**6, 9, 22**). (4) Notes in the middle ranges of the piano are more easily judged than the extremes (**3, 14**). (5) There is a tendency to judge low tones higher and high tones lower (**3, 14**). (6) Complex tones are more easily and accurately judged than pure tones (**1, 2, 3, 8, 9, 14, 21**). (7) Pitch-judging accuracy diminishes in the absence of practice (**6, 8, 9, 20**). (8) The most accurate judgments are those made quickly (**1, 2, 4, 6**).

A major inadequacy of the early studies was their lack of agreement concerning the boundary between absolute pitch and relative pitch. One semitone or larger unit was used in early studies as the criterion for absolute pitch. Recent investigators, notably Bachem (**1**) and Seashore (**17**), have set the criterion at a small fraction of a semitone. This lack of agreement concerning the nature of absolute pitch prevents reliable comparisons of data and accounts for many contradictory conclusions.

Procedure

Forty-two violinists with a median age of 19 years were chosen randomly from a list of violinists in the Stanford–Palo Alto area. Five were professionals; 14 were regular violin students; and 23 were occasional players. One-third of these violinists allegedly had absolute pitch. The experimenter tested all *Ss* to determine whether they conformed to the writer's definition of absolute pitch; namely, the accurate nam-

[1] The author gratefully acknowledges the assistance of Professor Paul W. Farnsworth.

ing of random tones with better than semitone accuracy.

The Conn Chromatic Stroboscope was used to determine the accuracy of the tones to .01 semitone. A violin *A*-string produced the *A*-440 stimulus tone. It is an open string tone; it is one of the most familiar tones to a violinist; and it is in the middle range, hence easier to judge than a high or low tone. At *A*-440 there is also a linear relationship between frequency as measured on the stroboscope, and pitch, which is the quality perceived by the subject (7).

Two laboratory periods were held at least 24 hr. apart. In one series the violin *A*-string was set to *E*, which had been determined by previous calibration on the peg. The following instructions were given each *S:*

> I will play a tone for 3 sec. You are to tell me if the tone I played was concert *A*-440, the pitch to which the violin *A*-string is normally tuned. If it is not *A* tell me how much to raise or lower it. I will then play the new tone. You are to tell me if that is *A* and so on. When you think I am playing *A*-440 say, "That is *A*." I will adjust the instrument as many times as you wish until your conception of *A*-440 is produced. There will be five judgments. Are there any questions?

The experimenter then played *E* below *A*-440 for approximately 2 sec. When *S* indicated his choice, the fork dial was adjusted until the *A*-disc on the stroboscope unit stopped. The reading was then taken from the fork unit dial. This procedure was repeated five times. The initial settings on the violin for each of the five judgments were *E*, $D\sharp$, $C\sharp$, *D*, and *E*. The progression upward to *A* was made by approximate half-steps which varied with instructions from *S*. The *S* was blindfolded during this series to restrict the cues to hearing alone.

Between the first and second series of tones, a 10-min. interview was utilized to gather background information and introspective data.

The second series was included to determine whether timbre cues would affect the pitch judgments. This series was played on the clarinet. The *S* was again blindfolded. He was told to instruct the experimenter to raise or lower the clarinet tones until *A*-440 was reached. The experimenter played five different series of clarinet tones. The order of tones varied according to *S*'s instructions. For example, if *S* said, "Play it up a perfect fifth," the intervening tones were skipped. . The interpolated tones were obtained by "lipping" the relatively fixed clarinet tones up or down as requested by *S*.

Wind pressure on the clarinet was kept constant for all *S*s in order to maintain equal loudness. Lewis (7) and Mursell (10) emphasize that loudness may be an influential variable in pitch-judging.

An interval of at least 24 hr. was planned between the preceding periods and the third series of judgments. The *S* was given the bow and violin with the *A*-string tuned to *E* below *A*-440. The other strings were randomly out of balance. The order of initial settings for the five judgments was *E*, $D\sharp$, $C\sharp$, *D*, and *E*. The following instructions were given to *S*:

> Tune the violin *A*-string to concert *A*-440 as closely as you can. You may tune all the strings if you wish, if you think it will help you get the *A*-string tuned more accurately. When you think you have played *A*, say "This is *A*," and continue playing that tone a few sec. longer so I can adjust the instrument. There will be five judgments. Are there any questions?

The *S* was not blindfolded during this series so that he could use visual, kinesthetic, and tactual cues in conjunction with his hearing.

Results and Conclusions

The data were treated by the *F* test of significance using the double-classification analysis of variance. When the *S*s tuned the violin, the average deviation from *A*-440 was 91 cents or almost 1 half tone. On the other hand, when the *S*s directed the *experimenter's* tuning, the average deviation from *A*-440 was 98 cents.

Some differences were apparent among the means of the judging tasks, but the significance test indicated that these differences could have been attributed to chance. In order to be significant at the .05 level, *F* should have been 3.11. The obtained *F* of 2.79 is significant at the .10 level of confidence. Since the differences fell in the direction which Cheslock (5) and Gough (6) predicted, it is possible that with more *S*s significant differences might have been found. The addition of visual, kinesthetic, and tactual cues, within the limiting conditions of this study, does not increase accuracy significantly.

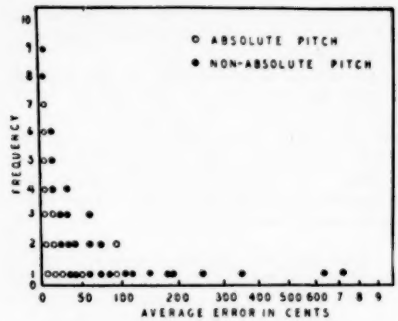

FIG. 1. Accuracy of absolute pitch *Ss* and nonabsolute pitch *Ss* in tuning a violin to *A*-440

This conclusion applies to *mean* differences. Some *individual* differences in judgment were very great. Baird (**3**), Whipple (**23**), and Riker (**14**) found that tone quality appeared to influence the accuracy of judgment. In the present study the unfamiliar timbre of the clarinet caused confusion in the judgments of several *Ss*. Nine expressed the opinion that they were completely "lost" on the clarinet tones but were quite confident of the violin judgments. The average deviation from the clarinet *A*-440 for all *Ss* was 123 cents, considerably over a semitone.

It was noted that when the violinists were asked to state whether they were more certain of their own tuning or more certain of their direction of the experimenter's tuning, a highly significant number were more certain of their own tuning than their direction of the experimenter's tuning.

In Fig. 1 is shown the accuracy of all 42 *Ss* on their own violin tuning. The cents scale is logarithmic. All the absolute pitch *Ss* were within 100 cents of *A*. Two nonabsolute pitch *Ss* were as accurate as the most accurate absolute pitch *Ss*. Forty-two per cent of the nonabsolute pitch *Ss* were within the range of the absolute pitch group. Errors of the nonabsolute pitch group ranged over an octave, whereas the errors of the absolute pitch group ranged over a whole tone only.

The above results are in fair agreement with Wedell's (**21**) finding that absolute pitch *Ss* have an average error of .5 of a semitone since the average for this study is approximately .25 of a semitone. The *Ss* in the present study were judging under more favorable conditions since they judged tones on their favored type of instrument and were aiming for the tone most familiar to them. Wedell's *Ss* judged piano and oscillator tones over a 24-note range.

Wedell (**21**), using 9 absolute pitch *Ss* and 11 nonabsolute pitch *Ss*, found a bimodal distribution of judgments clustering around .5 of a semitone and around 4 semitones away from the stimulus tone. In the present study there is no evidence of this type of distribution. All 42 *Ss* form a continuous distribution from almost perfect accuracy to an octave error. Riker (**14**), using musical and nonmusical *Ss*, also found ability to judge pitch continuously distributed.

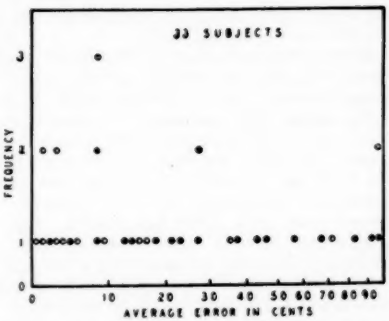

FIG. 2. Accuracy of absolute pitch *Ss* and nonabsolute pitch *Ss* in tuning a violin to *A*-440 through directions to *E* and while blindfolded (*Ss* with error > 100 cents eliminated)

The data in Fig. 1 indicate that the "absolute versus nonabsolute" pitch dichotomy has little meaning in describing fine pitch-judging accuracy. The overlap appears to be considerable. The 14 *absolute pitch* Ss had a mean deviation from *A* of 28 cents; whereas the *top 14 nonabsolute pitch* Ss had a mean deviation from *A* of 37 cents.

In Figs. 2, 3, and 4 are shown the mean deviations from *A*-440 of those Ss who deviated from *A* with an error less than 100 cents, and were directing the experimenter's tuning (Fig. 2) or making their own tuning (Fig. 3) on the violin, or were directing the experimenter's tuning on the clarinet.

The third problem of the study was, "What devices are used by the Ss to name tones accurately?" Some differences between the absolute and the nonabsolute pitch Ss are found here. The violinists were asked to describe how they knew when they played or heard *A*-440. This question was asked to uncover some of the cues employed to obtain more accurate judgments.

Of the 99 responses the most frequent was the expression of a vague feeling of disturbance and discomfort

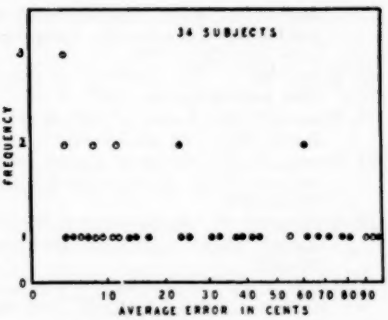

FIG. 3. Accuracy of absolute pitch Ss and nonabsolute pitch Ss in tuning a violin to *A*-440 directly (Ss with error > 100 cents eliminated)

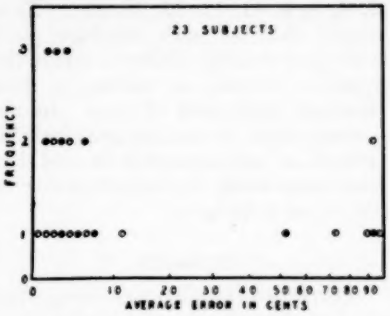

FIG. 4. Accuracy of absolute pitch Ss and nonabsolute pitch Ss in tuning a clarinet to *A*-440 through directions to *E* and while blindfolded (Ss with error greater than 100 cents eliminated)

when not precisely on *A*. This response was given by 36 per cent of absolute pitch Ss and by 24 per cent of the nonabsolute pitch Ss. More of the absolute pitch group gave responses which stated that the tones sounded too "brilliant" when sharp or too "dull" when flat.

The nonabsolute pitch Ss said that they used reference tones, visual imagery, auditory imagery, and kinesthetic cues more frequently than the absolute pitch group. Eight absolute pitch Ss admitted using other tones as reference points, but only after they had made their judgments; whereas the nonabsolute pitch Ss used the tones as direct reference points from which to find *A*.

Five responses which stated the use of laryngeal cues were given by nonabsolute pitch Ss. None of the absolute pitch group gave this response. Eight Ss claimed the use of visual and auditory imagery. *D* was the most commonly used reference tone utilized by the 12 Ss.

An item of interest is the group of 11 Ss who mentioned the unique quality of the notes. This finding has been noted by many investigators (2,

3, 4, 5, 9, 11, 12, 13, 17, 23). They found that *S*s with absolute pitch were particularly prone to stress the quality, chroma, or timbre in their absolute judgments of tone. In the present study 6 nonabsolute pitch *S*s, as well as 5 absolute pitch *S*s, said that they were using the unique quality of the notes in judging.

Summary

Forty-two violinists were selected to find *A*-440 under several conditions—when the violinist with all sensory cues at his command had the violin in his hands with the strings randomly out of balance; when *E* tuned the violin under the blindfolded violinist's directions; and when *E* tuned a *B*-flat clarinet to *A*-440 under the blindfolded violinist's direction. The data indicate that the addition of visual, kinesthetic, and tactile cues to hearing did not increase accuracy of tuning significantly. Violinists with alleged absolute pitch were more accurate than the nonabsolute pitch *S*s; but the great overlap questions the use of the term "absolute pitch" as a special indicator of accuracy. It was found, furthermore, that the nonabsolute pitch *S*s claimed to use reference tones, auditory imagery, and kinesthetic cues more frequently than the alleged absolute pitch *S*s. The more accurate the *S*, the less he seemed to depend upon these extraneous cues.

(Manuscript received May 8, 1950)

References

1. Bachem, A. Genesis of absolute pitch. *J. acoust. Soc. Amer.*, 1940, **11**, 434–439.
2. Bachem, A. Various types of absolute pitch. *J. acoust. Soc. Amer.*, 1938, **9**, 146–151.
3. Baird, J. W. Memory for absolute pitch. *Studies in psychology, Titchener commemorative volume.* Worcester: Wilson, 1917, 43–78.
4. Boggs, Lucinda, P. Studies in absolute pitch. *Amer. J. Psychol.*, 1907, **18**, 194–205.
5. Cheslock, L. Some notes on perfect pitch. *Amer. Merc.*, 1934, **32**, 459.
6. Gough, E. The effects of practice on judgments of absolute pitch. *Arch. Psychol.*, 1922, **47**, 1–93.
7. Lewis, D. Pitch, its definition and physical determinants. *U. Ia. Stud. Psychol. Music*, 1937, **4**, 346–373.
8. Meyer, M. Is the memory of absolute pitch capable of development by training? *Psychol. Rev.*, 1899, **6**, 514–516.
9. Mull, Gelen K. The acquisition of absolute pitch. *Amer. J. Psychol.*, 1925, **36**, 469–493.
10. Mursell, J. L. *The psychology of music.* New York: Norton, 1937.
11. Neu, D. M. A critical review of the literature on absolute pitch. *Psychol. Bull.*, 1947, **44**, 249–266.
12. Ogden, R. M. *Hearing.* New York: Harcourt Brace, 1924.
13. Petran, L. A. An experimental study of pitch recognition. *Psychol. Monogr.*, 1932, **42**, 1–132.
14. Riker, B. L. The ability to judge pitch. *J. exp. Psychol.*, 1946, **36**, 331–346.
15. Seashore, C. E. Acquired or absolute pitch. *Music Educators J.*, 1940, **26**, 18.
16. Seashore, C. E. Inheritance of musical talent. *Mus. Quart.*, 1920, **6**, 586–589.
17. Seashore, C. E. *Psychology of music.* New York: McGraw-Hill, 1938.
18. Seashore, C. E. *The psychology of musical talent.* New York: Silver Burdett, 1919.
19. Seashore, C. E. *The measurement of pitch intonation with the tonoscope in singing and playing.* Univ. of Iowa Series; Aims and Progress of Research II. Iowa City: Univ. Iowa Press, 1930.
20. Watt, H. J. *The psychology of sound.* New York: Macmillan, 1917.
21. Wedell, C. H. A study of absolute pitch. *Psychol. Bull.*, 1941, **38**, 547–548.
22. Wedell, C. H. Nature of absolute judgments of pitch. *J. exp. Psychol.*, 1934, **17**, 485–503.
23. Whipple, G. M. Studies in pitch discrimination. *Amer. J. Psychol.*, 1903, **14**, 289–309.
24. Wyatt, Ruth. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, **58**, No. 2.

BEHAVIOR STRENGTH AS A FUNCTION OF THE INTENSITY OF THE HUNGER DRIVE

BY GREGORY A. KIMBLE

Yale University[1]

The strength of the tendency to perform a learned reaction, according to current learning theory, depends upon the strength of motivation. The present study was designed to investigate the nature of the function which relates the latency of a conditioned operant response to the strength of the hunger drive. In the terminology of Hull (2) the problem is that of determining the relationship between response latency ($_{st}t_R$) and drive strength (D) when hours of food deprivation (h) and amount of food last eaten (A) are the determiners of drive. In the case of the hunger drive (D_H) it is often said that the intensity of this motive increases with the number of hours that the animal has been deprived of food ($D_H = f[h]$). But it is also possible to control hunger by varying the amount of food ingested, a technique employed by Skinner (8) and Reynolds (6). The latter investigator, for example, produced a low and a high drive state in two groups of animals by feeding the low-drive group 12 gm. of food 24 hr. prior to a learning experiment, and the high-drive group 3 gm. of food 24 hr. before the same experiment. The proper formula for the strength of the hunger drive, thus, may be $D_H = f(A,h)$.

Response latency should vary as a function of time since eating (amount last eaten constant) or as a function of amount eaten (time since eating constant) since either of these influences the intensity of hunger. Hunger in turn is a drive and determines, with habit strength ($_sH_R$), the value of an intervening variable, $_sE_R$, upon which response latency depends directly. Thus, even if the strength of habit is constant, latency should vary with drive, which partially determines the magnitude of $_sE_R$.

The relationship between response latency and hunger has not been traced adequately experimentally. Investigations of the relationship between behavior potentiality and the strength of hunger have presented results mainly in terms of numbers of trials to extinction and have reported latencies only incidentally. Such studies include those of Perin (5), Koch and Daniel (4), and Saltzman and Koch (7). The results of these investigations, as summarized by Saltzman and Koch, suggest that the curve relating trials to extinction to hours of food deprivation is one which rises rapidly during the first 2 hr., and then continues to rise at a much slower rate, at least up to 24 hr. If response latency reflects the same thing as number of trials to extinction, it is to be predicted that the *reciprocal* of reaction latency and number of trials to extinction will be related to drive strength in similar ways.

A special problem arising in connection with the functional relationship between hunger and its determining conditions, as pointed out by Zeaman and House (11), involves the concept of satiation. Hull's (2) earlier statement about behavior under satiation was based upon an extrapolation of the function obtained by Perin (5) to

[1] The experiment reported in this paper was performed in the psychological laboratory at Brown University.

zero hours, of food deprivation. His conclusion was that, at this point, resistance to extinction is about 28 per cent of its value at 24 hr. of deprivation. More recently Koch and Daniel (4) and Saltzman and Koch (7) have taken experimental issue with this conclusion. In an experiment similar to the Perin experiment they found that resistance to extinction was very low at low drive intensities. At zero hours of deprivation, the median number of trials to extinction was 1.0. Hull's (1) more recent account of the function which relates hunger to number of hours of food deprivation has been presented only in a very brief form. But apparently it is in line with the findings of Koch and his co-workers. Hull writes that "The functional relationship of drive (D) to one drive condition (food privation) is: from $h = 0$ to about 3 hours drive rises in an approximately linear manner until the function abruptly shifts to a near horizontal, then to a concave-upward course, gradually changing to a convex-upward curve reaching a maximum of 12.36σ at about $h = 59$, . . ." (1, p. 176). This statement is not, however, specific with respect to the characteristics of behavior under food satiation.

Purpose of the Experiment

The foregoing introduction poses three questions which identify the threefold purpose of this experiment: (1) What is the nature of the function relating behavior potentiality (reflected in reciprocal latency) to number of hours of food deprivation? (2) Are there changes in the strength of the tendency to perform a learned response correlated with the ingestion of food? (3) What are the characteristics of behavior acquired on the basis of the hunger drive when this drive is reduced to zero?

Experimental Procedure

Apparatus.—The response required of the *S*s in this experiment was a panel-pushing reaction similar to that previously employed by Webb (10). The apparatus was designed to measure the latency of this reaction as accurately as possible. This apparatus consisted of a box 11 in. long by 6¼ in. high by 5¼ in. wide with a masonite guillotine door in one end. When this door was raised, a black panel was presented. This panel, which was 4¾ in. high by 2¾ in. wide, was pivoted so that a very light pressure against it exposed a food cup and the food reward. A counterbalance kept the panel open after the response was made.

Raising the guillotine door 1 in. depressed a microswitch and started a Springfield timer. Opening the panel closed another switch improvised from a telegraph key, activated a holding relay which stopped the timer, and thus recorded the latency of the response to the nearest .01 sec. Except for the counterbalancing arrangement, the important details of the apparatus are shown in Fig. 1.

Subjects.—Twelve female rats from the inbred strain of Wistar stock maintained by the Department of Psychology at Brown University were the *S*s in the experiment. They were 73 to 75 days old at the beginning of training. One additional animal was discarded for failure to learn the panel-pushing reaction.

Scheduling.—The relationship between behavior potentiality and hunger is probably partly (and perhaps importantly) a function of the

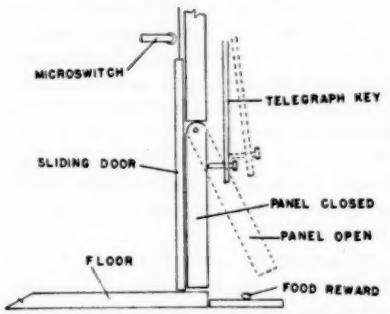

Fig. 1. Apparatus. This figure shows only the portion of the apparatus involved in the recording of the rat's response latency. When the sliding door is raised, the microswitch is depressed and a .01 sec. clock started. When the panel is pushed open, the telegraph key switch breaks the clock circuit by activating a holding relay.

maintenance schedule of the animal. For this reason, it appeared to be important to establish a constant motivational state under 24 hr. of deprivation before beginning the experiment. To this end, the animals were put on a feeding schedule for 11 days prior to the experiment. During this period, a record of the amount of food eaten in a 20-min. feeding period was taken on days 1, 3, 5, 7, 9, 10, and 11. The assumption was that a constant motivational state had been reached when the food intake became essentially constant from day to day. Eleven days on the 24-hr. deprivation schedule seemed to have accomplished this. On the ninth day, the animals ate slightly less than they had on the seventh day; and on the eleventh day, they ate an average of .2 gm. more than they had on the tenth day and 2.2 gm. more than on day 7. Food intake seemed to have stabilized at about 21 gm. of mash per day.

The *S*s were allowed to eat 20 min. per day on a mash composed by weight of 40 per cent ground Purina dog checkers and 60 per cent water. To this mixture a small amount of cod-liver oil had been added. The *S*s were fed on platforms near the experimental apparatus. On days when weight measures were taken, the feeding period was divided into four sessions. At the end of each 5 min., the food was removed and *S*s were weighed to determine their intake. This periodic interruption of eating is believed to be particularly effective as a preliminary to the use of a satiation technique, in that *S*s learn to eat rapidly. During the fourth 5-min. eating period they ate almost nothing.

Training.—After this scheduling procedure, *S*s were taught the panel-pushing habit in a single session involving 3 pretraining trials and 15 learning trials. On the pretraining trials, the panel was held open, and *S* was allowed to discover and to eat a small pellet of mash thus presented. On each of the 15 training trials, *E* waited until *S* was facing and had its forepaws in the half of the box containing the guillotine door. When this criterion position was assumed, the door was raised and the trial begun. Under this training procedure, learning is very rapid; response latency reached an asymptote on trial 11 at a median value of .4 sec.

Tests under varying numbers of hours of deprivation.—After preliminary training, the experiment proper was begun. In a random order on different days, *S*s were tested after 0, 2, 8, 15, and 24 hr. of food deprivation. Under each condition, *S*s were given five trials under the same conditions as training. Food was present in the food cup, and the same criterion based upon the animals' position in the box was used to determine when to raise the guillotine door and start a trial.

In all these conditions, except for the satiation condition, the reported number of hours of food deprivation is the length of time since the end of the last regularly scheduled 20-min. eating period. Because of certain precautions taken with the zero-hour condition, it requires additional comment. It seemed desirable to make every effort to produce complete satiation (if this is possible) and to obtain enough data upon each *S*'s behavior to form the basis for dependable statements. Three different procedures were employed for these two purposes. The first involved the treatment of the animals during the eating period. As already mentioned, most of the animals finished eating in 15 min. of their allotted 20 min. In an effort to encourage further eating, the food was removed from the feeding platforms after 20 min., stirred, and returned after 1 min. This occasionally produced more eating for a few seconds. When the animals had again ceased to eat, they were lifted and placed before the food cups. This never produced any additional eating, nor did attempts to feed them by spoon.

The second special procedure was designed to obtain more data under satiation than under the other motivational levels. For this purpose, the satiated animals were given four blocks (instead of the usual *one*) of five trials each on successive days. On these trials, an animal which did not respond to the panel within 5 min. of its exposure on any trial was removed from the apparatus and assigned an arbitrary latency of 5 min. for that trial and for the remaining trials in that five-trial series. Fortunately, this procedure was never necessary for as many as half the animals. Median latencies are, therefore, obtainable for this condition as well as for each of the others.

For the tests under satiation, all *S*s were fed at the same time. This meant that some *S*s were run under satiation later than others. This procedure, which looks more hazardous from the point of view of experimental design than it actually is, was employed for three reasons: (1) to aid in the mechanical details of running the experiment and keeping the *S*s on feeding schedule; (2) because previous observations upon the eating behavior of the *S*s had indicated that they would not eat for an hour or more after the satiation procedure used in this experiment; and (3) because this procedure provided latency measures at several points during the period immediately after eating. Of these three considerations, the last was the most important.

Specifically, *S*s tested under satiation could be regarded as being run under different (short) periods of deprivation ranging from 10 min. to 60 min. These values were arrived at by dividing the 12 *S*s into four groups of three *S*s determined by the order in which they were run.

Since the order of testing under satiation was randomly determined from day to day, it seemed proper to combine data for these days, thus providing 60 latencies (3 animals × 5 trials × 4 days) upon which to compute a median latency. By averaging the times after satiation at which each of these subgroups was run, and by rounding to the nearest multiple of 5, we thus obtain latency measures corresponding roughly to 10 min., 30 min., 50 min., and 60 min. deprivation.

As a third special procedure employed in the tests under satiation, a check on the possibility that the animals might not have been completely satiated in spite of our precautions was attempted on a fifth day. They were fed according to the usual satiation procedure. When *S* was satiated, it was given a single trial in the panel-pushing apparatus. But in this case, one of the food cups was placed in the apparatus to provide an easily accessible supply of food. If *S* ate this food (3 *S*s), it was removed from the apparatus to the feeding platform, allowed to eat, and run again later. If it did not eat (9 *S*s), the single test-trial was administered immediately.

Tests after various amounts eaten.—Two further sets of measurements were made on the same *S*s. In this phase of the experiment, the *S*s were interrupted in their eating and run within 1 min. after this interruption. The conditions used were one in which the *S*s had been allowed to eat for 10 min., and another in which they had been allowed to eat for 15 min. These two conditions, together with the 24-hr. deprivation condition and the satiation condition, provide a sample of behavior after various amounts of prefeeding.

Summary of experimental conditions.—The experiment provides data on the behavior of the *S*s under 11 different motivational conditions. There are three conditions in which the drive-controlling variable is the amount of time allowed for eating after 24 hr. of deprivation. These times are 0 min., 10 min., and 15 min. In the other eight conditions drive strength is determined by the time since last eating to satiation. These times are 10 min., 30 min., 50 min., 60 min., 2 hr., 8 hr., 15 hr., and 24 hr. For each condition, the dependent variable is the latency of the conditioned reaction.

Results

The major findings of the experiment are presented in Fig. 2, where the reciprocal of the median latency is plotted as a function of time either

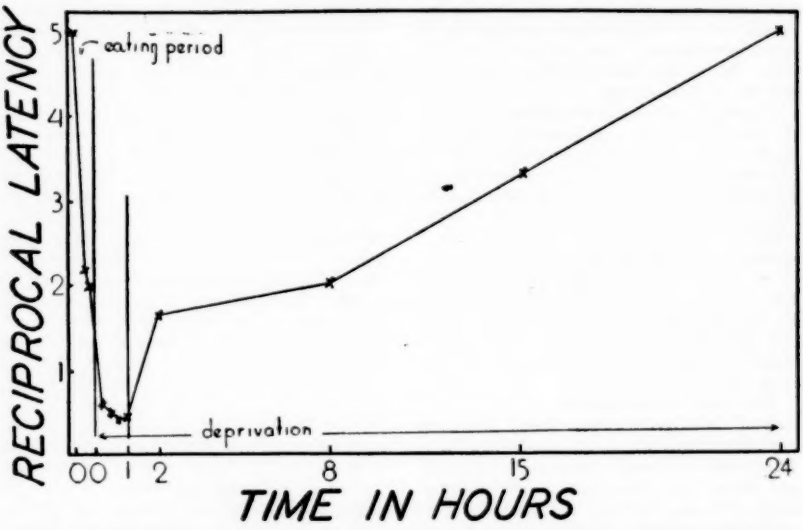

Fig. 2. Median reciprocal response latency as a function of time allowed for eating and time of deprivation. The left-hand portion of the graph represents changes in behavior strength associated with ingestion of food. The second and third sections present reciprocal latency as a function of number of hours of food deprivation.

since the beginning of eating or since the termination of eating, depending upon the section of the graph. Median latencies are used because of the occasional indeterminate latencies under the low-drive conditions which make the use of other averages inappropriate. The reciprocal of the median latency is used for purely aesthetic reasons. This measure increases where we would want to consider drive as increasing and decreases where we would want to consider drive as decreasing.

Figure 2 is divided into three sections. The portion of the figure at the left shows latency changes which accompany the ingestion of food. In terms of time allowed for eating prior to their determination, the three points represent 0 min., 10 min., and 15 min. The first of these points (0 min.) is, of course, the same value as that plotted for 24-hr. deprivation. The expected rapid decrease in reciprocal latency as a function of time allowed for eating is readily seen.

Orthodox statistical tests are not applicable to the data of this experiment for three reasons: (1) Variability under low drive is many times as great as under high drive; (2) latency distributions under low drive are very badly skewed; and (3) when a criterion of no response is used, indeterminate latencies are sometimes obtained. For these reasons, the results of the experiment have been evaluated statistically by means of the nonparametric sign test described by Dixon and Mood (3). The sign test is a special application of the binomial theorem in which one merely counts the number of positive and negative differences obtained when one subtracts measures under one condition from those under another and determines the significance of the departure from a 50-50 distribution. For example, in this experiment, when one subtracts the median latency produced in the 10-min. prefeeding condition from the same value for the 15-min. prefeeding condition for each animal, eight positive differences are obtained. Using a single-tailed test (since we have made predictions of direction) this result is significant at only the .20 level of confidence. By the same test the difference between the median latencies obtained under the 24-hr. deprivation condition and either of the above prefeeding conditions is significant beyond the .01 level of confidence. Similarly, the difference between median latencies for either prefeeding condition and the satiation condition is significant beyond the .01 level.

The second, intermediate, portion of the graph shows the changes in reciprocal latency which occur in the first hour after satiation. These points are considered separately because the values they represent were determined in a manner somewhat different from that used in the determination of the points plotted in the other portions of the graph. The most striking thing suggested by the trend apparent in these points is that response strength may continue to decrease for the first hour following satiation. This finding is not unique. Saltzman and Koch (7) showed a similar effect for the first 30 min. after satiation. Thus, this finding confirms their result and suggests that it persists for 1 hr. or longer.

The reliability of this trend was also evaluated by the sign test. The method was to compare the median response latency of each *S* when it was run in the first block of *S*s, with the same measure obtained on another day when the *S* was run in the last squad of three animals. As a result of the random assignment of order, it happened that only nine of the *S*s supplied data in both conditions. Of these, seven produced longer latencies in the second condition. The probability of such an outcome on the basis of chance is 46 in 512, a result significant at the .10 level of confidence.

Thus we have obtained no evidence for an increase in behavior strength during the first hour of deprivation. And, while the evidence for a decrease is not impressive, it does permit us to suggest this as a possibility. Between 1 and 2 hr. of deprivation, a sharp increase in behavior strength occurs. Beyond 2 hr., the function is one which increases more slowly up to the next regularly scheduled feeding period. The initial increase in behavior

strength occurred in all Ss, making a statement of statistical dependability superfluous. The difference between the 2- and 8-hr. performance is not statistically reliable. By the sign test the difference between the 8- and 15-hr. results is significant at the .01 level; that between the 15- and 24-hr. conditions is significant at about the .10 level. An examination of the individual latency distributions indicates a possible explanation for this latter result, in that some Ss were apparently responding with maximal speed under the drive produced by 15-hr. food deprivation.

A final set of observations obtained in this investigation was that during the five-trial tests under satiation, the response latency becomes progressively longer. Median latencies for the five trials are .9, 12.0, 37.3, 123.5, and 145.9 sec. That this trend is reliable is indicated by the fact that when median latencies, based on all trials under satiation, are examined S by S, there is a consistent tendency for this measure to be larger on the last two than on the first two trials. To make this comparison, two Ss were disregarded for producing so many indeterminate latencies as to make it impossible to determine the trend. Of the ten remaining Ss, nine show the effect, a finding significant at about the .01 level of confidence. The elimination of the two Ss for whom median latencies could not be determined also has the effect of demonstrating that this increase in latency is not merely an artifact of the procedure of assigning latency values of 5 min. for all trials beyond the one on which the first such latency appears. For the ten Ss remaining, the trend being evaluated was not affected by the presence or absence of such indeterminate latencies.

Discussion

The experiment just reported was designed to provide information on behavior strength as a function of drive intensity when the drive-controlling operation is either prefeeding or deprivation.

Associated with prefeeding, it was found that there occurs a rapid drop in the reciprocal latency of a conditioned response based upon food reward. This finding is easily predicted upon the basis of any of the current learning theories. Hull's (2) theory, which makes the matter most specific, would hold that hunger (D) decreases with the ingestion of food, thus decreasing the magnitude of an intervening variable, ${}_{S}E_{R}$. This latter decrease is reflected directly in an increased latency or a decreased reciprocal latency. The only importance which attaches to our findings is that the predicted has been demonstrated in a new experimental situation.

Somewhat more surprising is the finding that there is no increase (perhaps some decrease) in behavior strength for about 1 hr. after eating to satiation, when satiation is defined as refusal to eat. This finding confirms that of Saltzman and Koch (7). To the physiologically minded psychologist it should certainly suggest that satiation depends in part upon bodily absorption of food products.

For the period from 1 hr. to 24 hr. after satiation, the curve relating reciprocal latency to the number of hours of food deprivation is one which rises rapidly to 2 hr., and then more slowly to 24 hr. of deprivation. In its gross aspect, this curve is much like that presented by Saltzman and Koch (7) and also resembles that described by Hull (1). It, therefore, appears that the function relating the strength of

the hunger drive to food deprivation is not restricted to a single apparatus, or a single response measure, or a single strain of rats. The most important question raised by such functions concerns the sudden shift in acceleration which appears at 2 or 3 hr. of deprivation. One possibility is that this represents a residual effect of the rat's 2-4 hr. ad lib. feeding schedule.

Of special interest is the behavior observed under satiation. In addition to the expected protracted latencies, it was noted that these latencies increased regularly through a five-trial test series run under satiation. This finding appears to be more difficult to explain than any of our other results. The very long latencies on trials 4 and 5 present no difficulty. They may be interpreted as simply reflecting the reduction of drive and, therefore, of ${}_{S}E_{R}$, to nearly zero. It is the short latencies on the early trials which present a problem. One first suspects incomplete satiation. This explanation, however, cannot account for the behavior observed on the fifth day of testing under satiation, when food was available in the apparatus. Under this condition, the available food was never eaten; but all *S*s responded to the panel. The median latency of these responses was .2 sec. Thus, when there is an additional criterion in terms of which to justify the statement that the *S*s were satiated, it is found that they make at least one short-latency response. In the light of this evidence, one is forced to accept the view that some motive other than hunger in the usual sense was energizing this behavior, and (more important) that this motive functions in the absence of the original motive under which the habit was learned. One possibility is the operation of an acquired drive of the sort postulated by Tolman (**9**), whose general position is that techniques or responses which have, in the past, led to primary reinforcement somehow become ends in themselves. However, the acceptance of this position raises at least two questions: (1) Why should this acquired drive "extinguish" as rapidly as the latency changes observed in the trials under satiation lead one to suppose it does? (2) In about 50 per cent of the responses obtained under satiation, the *S*s ate the available food, even though they just refused the same kind of food in a familiar container in the apparatus. How shall this be explained? One suggestion is that behavior which has previously terminated in eating tends to complete itself; i.e., to continue to terminate in eating, even when the animal is satiated. But this is no explanation—rather, it is a restatement of an observation.

Fortunately, these questions are amenable to experimentation. If the behavior observed under satiation is mediated by a secondary drive, it should be possible to increase the strength of a response tendency tested under satiation by increasing the number of trials for the acquisition of this drive. And, if behavior which has, in the past, terminated in eating tends to complete itself under satiation, then the presence or absence of food should influence performance even under satiation when the food presumably has no drive-reduction value. One would predict a greater persistence of the response tendency under satiation for conditions where the food is present than for conditions where the food is absent.

Summary

Twelve female albino rats were taught a panel-pushing response to

obtain food after 24 hr. of food deprivation. Then they were tested under different amounts of drive when the drive-controlling operations were prefeeding or deprivation. Measurements of behavior strength in terms of median reciprocal latency of the response were obtained after 0, 10, and 15 min. of prefeeding, and after deprivations of 10 min., 30 min., 50 min., 60 min., 2 hr., 8 hr., 15 hr., and 24 hr. Associated with prefeeding there occurred a rapid decrease in reciprocal latency which persisted for the first hour after eating to satiation. Between 1 hr. and 2 hr. of deprivation a rapid increase in reciprocal latency occurred. Beyond 2 hr. this measure increased more slowly up to 24 hr. The shape of the function resembles that presented by Saltzman and Koch (**7**) for resistance to extinction. It also is like a curve described by Hull (**1**). The characteristics of the behavior observed under satiation suggest the operation of a motive other than hunger, perhaps an acquired drive of the sort postulated by Tolman (**9**).

(Manuscript received May 2, 1950)

References

1. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton Century, 1943.
3. Dixon, W. J., and Mood, A. M. The statistical sign test. *J. Amer. statist. Ass.*, 1946, **41**, 557–566.
4. Koch, S., and Daniel, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*,1945, **35**, 167–185.
5. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 95–109.
6. Reynolds, B. The relationship between the strength of a habit and the degree of drive present during acquisition. *J. exp. Psychol.*, 1949, **39**, 296–305.
7. Saltzman, I., and Koch, S. The effect of low intensities of hunger on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1948, **38**, 347–370.
8. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton Century, 1938.
9. Tolman, E. C. *Drives toward war.* New York: D. Appleton Century, 1942.
10. Webb, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.
11. Zeaman, D., and House, B. J. Response latency at zero drive after varying numbers of reinforcements. *J. exp. Psychol.*, 1950, **40**, 570–583.

A VERIFICATION OF THE GUESSING-SEQUENCE HYPOTHESIS ABOUT SPREAD OF EFFECT

BY C. A. FAGAN AND A. J. NORTH

Southern Methodist University

In 1933 Thorndike (**4**) reported the spread-of-effect phenomenon, advancing the hypothesis that reward strengthened not only a rewarded stimulus-response connection, but, also, and to a lesser degree, adjacent connections. Since then a number of studies (**1**, **2**, **3**) have tended to show that the "after" gradient (the only one reliably demonstrated) may be attributed to nonrandom *guessing habits*. That is, a tendency to prefer certain sequences of responses over others would of itself provide for a gradient of repetition after fixated rewarded responses. The present study is based on an experiment by Sheffield (**3**), in which "subjects were asked to try to divine—through clairvoyance—the sequence of random numbers in four specified columns of Tippett's tables." In analyzing the data, Sheffield looked only for an "after" gradient. A "repeat" was defined as the repetition of the same guess in the same ordinal position in two successive columns. A reliable "after" gradient, not attributable to reward or punishment but, rather, to guessing habits, was obtained.

The present study was undertaken in order to verify the guessing-sequence hypothesis about spread of effect by duplicating Sheffield's experiment with certain additional controls and modifications. Sheffield allowed his *Ss* to write their guesses in four columns. Possibly the sight of what had just been written might have influenced subsequent guesses. In the present study, *Ss* gave their responses orally, so that they could not refer to previous guesses. Sheffield's *Ss*, furthermore, simply wrote down a series of numerical guesses, but not in response to designated stimuli. In the present study the adequacy of the guessing-sequence hypothesis was tested by having *Ss* guess numbers which had been paired with verbal stimuli.

METHOD

Twenty-four university students (volunteers) served as *Ss*. The materials consisted of a list of 50 adjectives and sheets of paper for tabulating *S*'s guesses. The *Ss*, who were tested individually, were instructed as follows:

> This is an experiment in extrasensory perception. I have a list of 50 words which have been paired with numbers from 1 through 10. To insure random pairing of the numbers with the words, I put 10 cards, each bearing a number from 1 through 10, in a box. I shook the box and drew a number. I wrote this number beside the first word in the list and put the card back in the box. I shook the box and drew a card and put the number down beside the next word. I continued this process until all 50 words had been paired with a number.
>
> This is what I want you to do: I will call out a word and I want you to respond with any number from 1 through 10. We will do this for all 50 of the words. I will go through the list 4 times.
>
> Don't attempt to follow or devise a pattern, but guess in a random fashion. If you possess extrasensory perception, it will show up in guessing rather than in adherence to any system.
>
> Remember, the numbers range from 1 through 10; guess randomly.
>
> Do you have any questions? Ready!

After completion of the first condition of the experiment, these additional instructions were given:

> The second part of this experiment concerns numbers only. Using the method I described before, I devised a list of 50 numbers ranging from 1 through 10. When I say "Go," I want you to begin calling num-

bers from 1 through 10 in an attempt to reproduce this list. Give any of these numbers that comes to mind, as there is no order or system to the list. You will be given 4 trials. I will stop you at the end of each list and tell you when to begin the next. Try to give the numbers at a fairly rapid, even pace. Like this: 2 . . . 1 . . . 7 . . . 3 . . .

Remember, the numbers range from 1 through 10; guess randomly.

Do you have any questions? Ready? Go!

In order to counterbalance any positional effects, half the *Ss* were presented with the "words-numbers" list first, and half began with the "numbers-only" list. The instructions were modified to fit the particular condition being presented first.

Results

Repetitions were classified and scored as in Sheffield's (3) study. Each repetition of the same guess in the same ordinal position in two successive series of 50 responses was scored as a "repeat." For each *S* the number of chance repeats and the number of repeats in each of the five positions following chance repeats was determined. Next for each *S* the percentage of repetition for each position following chance repeats was determined. These distributions of percentages, however, tended to be positively skewed. Therefore, the median rather than the mean was chosen as the measure of central tendency. The median per cent was determined for each position in order to discover whether an "after" gradient of repetition existed. The results are shown in Table I. It will be noted that in each condition a "chance" level of repetition (assumed to be 1 in 10, or 10 per cent) was approximated by about the third position.

A problem arose in examining the significance of the apparent decreasing gradient of repetition. The general method adopted was to compare the frequency of repetition in the first position with the frequency in each of the other four positions. When, however, the distribution of differences in percentage of repetition between, say, position 1 and position 2 was examined, the data were found to be too variable for purposes of a sensitive test of statistical significance. Many of the large differences, for example, were due to the presence of *Ss* who had made few chance repeats, so that a small difference in the frequency of repeats would be magnified when expressed as a difference in percentage. Furthermore, there was a tendency toward skewness of the distribution of differences in percentages, thus rendering somewhat questionable the application of tests of significance assuming normality. The distributions of differences in *frequencies* of repetition, however, were *not* subject to the foregoing objections. The logic of the test of significance was as follows: If there were no downward trend in the frequency of repetition from position 1 to, say, position 2, then the mean of the distribution of differences in frequency ought to equal zero. The *t* test of significance was applied and evaluated for 23 degrees of freedom. In both the "words-numbers" and "numbers-only" conditions, there were significantly (at better than the .05 level of confidence) fewer repetitions in positions 2, 3, 4, and 5 than in position 1. Thus, there was a decreasing gradient of repetition.

TABLE I

Gradient of Repetition Following Chance Repeats Based on Median Percentage of Repetition

Condition	Position after Chance Repeat				
	1	2	3	4	5
Words-numbers	16.0	12.1	9.8	9.8	10.2
Numbers only	17.9	15.0	10.8	10.6	11.4

Summary

In order to determine if guessing habits could account for the spread-of-effect phenomenon, and in an attempt to verify results obtained by Sheffield, 24 *Ss* guessed numbers in the two conditions of this study. No reward or punishment was involved in either condition. Under the "numbers-only" condition, *Ss* were told that a list of 50 numbers from 1 to 10 had been constructed. Under the "words-numbers" condition, *Ss* were told that a list of 50 words had been paired with numbers from 1 to 10. As *E* read the words, *S* guessed orally the corresponding numbers. Four trials were given under each condition.

A declining gradient of repetition was found under both conditions of the experiment. The results follow those obtained by Sheffield and support the possibility that spread of effect may be accounted for in terms of guessing habits.

(Manuscript received May 15, 1950)

References

1. Jenkins, W. O., and Cunningham, Leta M. The guessing-sequence hypothesis, the 'spread of effect,' and number-guessing habits. *J. exp. Psychol.*, 1949, **39**, 158–168.
2. Jenkins, W. O., and Sheffield, F. D. Rehearsal and guessing habits as sources of the 'spread of effect.' *J. exp. Psychol.*, 1946, **36**, 316–330.
3. Sheffield, F. D. 'Spread of effect' without reward or learning. *J. exp. Psychol.*, 1949, **39**, 575–579.
4. Thorndike, E. L. A proof of the law of effect. *Science*, 1933, **77**, 173–175.

CONDITIONING IMAGERY

BY CLARENCE LEUBA AND RALPH DUNLAP

Antioch College

Classical conditioning has been motor conditioning, involving simple muscular or glandular responses. It has been produced mainly in animals and in infants. Attempts to condition more complex reactions and in adults have met only with indifferent success. The notion has accordingly developed that, as a principle of learning, conditioning could only account for certain limited, simple forms of learning. It is generally considered inadequate to explain learning involving total situations, complex patterns of response, and especially the numerous symbols which loom so large in the thoughtful learning of the adult or older child.

It is the purpose of this paper to describe the procedures whereby a type of symbol, namely imagery, can function as a conditioned stimulus in adults. This is an extension of what has come to be known as sensory conditioning.

Previous experiments have shown that sensory conditioning can be produced by using the same general procedures as in classical motor conditioning (**1**, **2**). For instance, an *S* will report the odor of creosote when a bell is rung, if the bell has been rung a number of times while *S* was actually smelling creosote. The present experiment was designed to determine whether such conditioned sensations could be elicited by imagery of the conditioned stimulus, i.e., by merely imagining the sound of the bell.

In some of the sensory conditioning experiments (**1**), hypnosis had been found useful for excluding irrelevant variables, for securing concentration on the conditioned and unconditioned stimuli, and for producing amnesia for the conditioning session. It was used in this experiment for these same purposes.

Method

Four *S*s, after having shown the phenomena typical of deep hypnosis, were put through the following procedures: (1) a conditioning session under hypnosis in which *S* concentrated on two sensory phenomena, such as the sound of a bell and a pin prick to the hand; (2) awakening from hypnosis, after amnesia had been created for the conditioning session; (3) while in the waking state being asked to imagine a number of situations, used as controls, and to report any experiences which might occur in response to them; and (4) being asked to imagine the conditioned stimulus—such as the sound of the bell—and to report any experience appearing in response to it.

The *S*s were college students who responded to a request for persons interested in being hypnotized and in whom deep hypnosis was readily produced.

Results

For all four *S*s and in each of a variety of situations, conditioned sensations (imagery) usually appeared in response to imagining the situation to which those sensations had originally been conditioned. This occurred even though *S* had no memory of the conditioning session and was, therefore, much puzzled by suddenly experiencing "shooting pains" in the hand, for example, when imagining that a bell was being rung.

Protocols: J. L.

When J. L. gave evidence of being deeply hypnotized, her attention was directed to sensations in her right hand and to sounds. A small doorbell was sounded, and simultaneously *E* pricked *S*'s right hand with an algesiometer several times. *S* was asked what had occurred and replied: "I heard a doorbell and a pin pricked my

hand." The stimuli were repeated three times. *S* was awakened following suggestions that she would be amnesic for the hypnotic period. When questioned later, *S* reported that she could remember nothing except going to "sleep." *S* was then asked to imagine experiencing the stimuli which *E* read to her and to report any ideas or experiences which might accompany the imaginings. She was told first of all to imagine as vividly as possible that she was smelling gasoline. Her replies to this, and the other imagined situations were as follows:

Smelling gasoline—Well, I like that smell.
Seeing a green light—You mean a traffic light? Emerald green? Well, I think of how a dark room with a green light looks very eerie.
Hearing a doorbell—[puzzled expression] When I imagine the doorbell ringing, I get a pain; a sharp pain that goes through me [*E:* Where?] In my right hand [rubs area previously pricked by algesiometer]. It's a sharp movement, like a needle; it comes and goes.
Hearing a train whistle—I think of it when I am in bed, not asleep but just about to go asleep; and how it sounds far off but really isn't.

Protocols: J. M.

With J. M. two conditionings were attempted in one hypnotic period. First a mixture of creosote and oil of cloves was held under *S*'s nose while *E* snapped his fingers loudly. *S* was asked to identify the stimuli. The second conditioning employed the same stimuli as for subject J. L. and proceeded similarly, except that J. M. identified the bell as an alarm clock going off rather than a doorbell. After the hypnotic period, *S* was asked to report whatever he experienced while imagining certain situations:

Seeing a green light—I think of envy.
Tasting something salty—I think of being at the ocean—the mist.
Hearing someone snap his fingers—[shook his head, said nothing].
Smelling gasoline—Fire, I think of fire.
Hearing an alarm clock go off—[surprised expression] I felt a sharp pain in my left hand [rubs it].

The *S* was asked to imagine each of the situations again, increasing the vividness of the imagery if he could. *S* replied that he thought he wasn't very good at imagining and that it was difficult for him to imagine these things. On this second time through the list he was asked specifically after each situation about any experiences he might have. None of the situations awakened any experiences. *S* said he could not imagine the situations at all vividly; he said the image of the alarm clock going off was brief and that he could not imagine it at all the second time through the list.

Protocols: J. J.

In the first session with J. J., he was pretested for his responses to the imagined situations. The situations and his responses to them were as follows:

Smelling gasoline—burning flames.
Seeing a green light—I keep seeing white through the green.
Hearing a doorbell—someone answering the buzzer or doorbell.
Hearing the radio receiver hum—a record.

After deep hypnosis had been induced, the doorbell and the pin prick to the hand were paired three times. When *S* was asked what he had heard, he called the doorbell a buzzer. Upon being awakened he was again asked for his associations to the imagined situations for which he had been pretested. He now gave the following responses:

Gasoline—driving an automobile.
Green light—[*S* closed his eyes] a dark room.
Buzzer—Feels sort of like being stuck by a pin, can't locate it particularly. Funny idea. [*S* rubbed his hand; then looked at it.]
Radio hum—hear a record.

In the second session, J. J. was pretested for his responses to the following imagined situations:

Hearing someone blow a whistle—heart jumps; scares me.
Having someone rub your arm—[no response].
Tasting salt—water.
Seeing a red piece of paper—[closes eyes; no response].

The situation used in the conditioning under hypnosis was smelling oil of cloves while *E* rubbed *S*'s arm just below the shoulder. These stimuli were paired for less than a minute. *S* reported that the arm-rubbing alone was then sufficient to cause him to smell the odor of oil of cloves. After the hypnosis had been terminated and *S* was amnesic for the hypnotic period, he responded as follows to the list of imagined situations:

Whistle—train.
Arm being rubbed—[Sniffs.] Some kind of chewing gum, Beechnut or spearmint—no, not spearmint. A strong smell, more like cloves. [*S* wrinkled his nose.]
Salt—feels dry.
Red paper—sun.

A third experimental session with J. J. followed immediately the one just reported. *S* was rehypnotized and instructed to hear and respond to a new *E*. This *E* gave *S* a bottle of camphor to smell. *S* reported that it smelled like Vicks (presumably the medicament, Vicks Vapo-Rub). *S* was then asked to open his eyes and was given a white card on which a simple geometric design had been drawn with a red crayon (a small square enclosed by a slightly larger one which touched the small square only at the latter's corners). *S* was told to gaze intently at the design while sniffing the camphor. After about a minute of this, *S* was told to close his eyes, the camphor bottle was removed, and hypnosis was continued for a short time. *S* was then awakened, amnesic for the hypnotic period. After a brief conversation, *S* was given a blank card, requested to look at it, and to report if anything should appear on it. He was then requested to imagine the following situations one after another: hearing the tolling of a bell, having someone rub his arm just below the shoulder, smelling Vicks Vapo-Rub, drinking coffee, and eating roast beef. Only when he was imagining himself smelling Vicks Vapo-Rub did something appear on the card. To this olfactory imagery, he responded as follows: "This sounds silly, but there is something there all right; some geometric form, like a triangle. It's red and a square or a triangle. [*E:* Is it like a circle?] No, more like a . . . maybe I could draw it. [On a blank card, *S* drew two figures.] No, that's not right." *S* then drew an almost exact replica of the figure which had been used in the conditioning process. He seemed satisfied with his second drawing except that he thought that perhaps the small square might extend beyond the diamond-shaped one, but eventually decided that it did not.

When J. J. imagined having his arm rubbed just below the shoulder, he did not see anything on the card, but he did wrinkle his nose and rub it. This response apparently had not been completely extinguished in the previous session.

Protocols: H. W.

After H. W. had been deeply hypnotized, he was shown a 3 by 5 in. white card on which a small diamond with the letter *S* in the center had been drawn with a red crayon. Simultaneously, a tin snapper was sounded for approximately 30 sec. Upon being awakened amnesic for the hypnotic period, a brief conversation ensued after which *S* was given a blank white card and asked to look at it and to report if anything should appear on it as he imagined various things: (1) eating and tasting roast beef, (2) smelling ammonia, (3) hearing a metal snapper, (4) feeling velvet cloth, and (5) smoking a cigaret. *S* did not report anything on the card, except when imagining the sound of a snapper. At that point, he smiled, looked intently, and reported seeing a "diamond." *E* inquired: "Any color?" *S* replied: "Yes, red; and there's a red *S* in the middle of the diamond." *S* then turned the card over. "It is there too. It's vague now and disappears." After the design had disappeared, *E* actually snapped the snapper a number of times. *S* smiled, looked more intently at the card, and said, "It's there when you snap the snapper." He examined the card carefully and said, "What goes on here?"

In most cases the *S*s seemed genuinely surprised and puzzled by the conditioned sensations. The latter appeared suddenly and as if out of the blue. They seemed incongruous. It just didn't make sense to J. J., for instance, that a geometrical design should appear on a blank card whenever he thought of the odor of Vicks Vapo-Rub. He examined the card carefully for a concealed drawing on it. It seemed equally preposterous to J. L. that whenever he thought of a doorbell, he got shooting pain in his right arm.

That the *S*s were merely trying to do what they thought the hypnotist expected them to do, is contradicted by this automatic and involuntary character of the conditioned sensations. It is contradicted also by the following facts: The *S*s were unaware of the possibility of sensory conditioning and that the experiment dealt with it; in some cases, the *S*s did not at first consciously associate the conditioned sensation with what they had just been asked to imagine; the *S*s were in a normal nonhypnotized state when the tests were run; and it did

not make any difference to have the tests run by an *E* other than the hypnotist.

Conclusions

In this experiment we have shown that imagery may occur in response to imagining the stimuli to which it has previously been conditioned. We used only simple sensory phenomena, such as the sound of a bell or a pin prick. It remains to be seen whether the imagination of more complex patterns can also become determiners of complex experiences through a typical conditioning process.

(Manuscript received May 20, 1950)

References

1. Leuba, C. Images as conditioned sensations. *J. exp. Psychol.*, 1940, **26**, 345–351.
2. Ellson, D. Hallucinations produced by sensory conditioning. *J. exp. Psychol.*, 1941, **28**, 1–20.

FACILITATION AND INTERFERENCE IN PERFORMANCE ON THE MODIFIED MASHBURN APPARATUS: II. THE EFFECTS OF VARYING THE AMOUNT OF INTERPOLATED LEARNING[1]

BY DOROTHY E. McALLISTER AND DON LEWIS[2]

State University of Iowa

The first paper of this series (2) was concerned primarily with losses in performance on the *standard* Mashburn task during the early stages of relearning (RL) when a constant amount of interpolated learning (IL) of the *reversed* task was given subsequent to different amounts of original learning (OL). Amounts of learning were specified in terms of number of practice trials. Results for experimental and control groups showed clearly that significant amounts of of both facilitation and interference appeared during IL, that the amount of retroactive interference was dependent on the amount of OL, and that acquired skill in performing on the standard task changed very little, if at all, over periods of several days without practice. As a consequence of the latter finding, a separate control group was not needed for each experimental group.

For the analysis of the effects of different amounts of OL and IL, 12 experimental groups were used. After 10, 30, or 50 trials of OL on the *standard* Mashburn task, *S*s received 10, 20, 30, or 50 trials of IL on the *reversed* task, and then relearned the standard task for 20 trials. All trials were given at the rate of 10 per day, and each trial was for 2 min., with 15-sec. intertrial rest periods except for a 2-min. rest between trials 5 and 6. Practice occurred on consecutive days (including Sundays). On each trial *S* made as many matches of response and stimulus lights as possible. In addition to a matches score, each *S* made an error score. An error was counted whenever *S*, immediately after completing one match, moved a control in a direction opposite to that required for completing the next match. Three errors were possible on each match, one on each control (1).

The data used in comparing the performances of the various experimental groups consisted of trial-by-trial means of number of matches and means of total number of errors. A decrease in the number of matches from the last trial of OL to the first trial of RL has been taken as a sign of interference. An increase in errors from the last OL trial to the first RL trial also signifies interference. Facilitation in performance is indicated by increases in number of matches and/or decreases in number of errors.

Effects of Varying the Amount of IL

Analysis of the facilitating and interfering effects of four different amounts of IL preceded by one of three amounts of OL requires that the 12 experimental groups be arranged into the three subexperiments indicated in the first column of Table I. With two exceptions, the 12 groups were the same as those used in studying the effects of different amounts of OL (2, Table I). The exceptions were Groups 29B and 31A. Groups 29 and 31, which are not included here, failed to reach levels of performance during OL which were comparable to the levels

[1] This work was done under Contracts N5ori-57 and N9onr-93801 between the State University of Iowa and the Special Devices Center, Office of Naval Research.

[2] Alfred H. Shephard, Walter Spieth, and James L. Hedlund gave valuable assistance in running the *S*s and analyzing the data.

attained by the other groups within their subexperiments. These differences were statistically significant. It was necessary, therefore, to replace Groups 29 and 31 by new Groups 29B and 31A.

Because the four groups in each of the three subexperiments did not perform identically during OL, it was necessary to evaluate the differences that appeared. For this purpose, analysis of variance techniques (**3**) were employed. Simple variance analysis was applied separately in relation to the individual matches scores and individual error scores for the four groups in each subexperiment. For Subexperiment E, tests were made on the scores on trial 10 of OL and on trial 1 of IL; for Subexperiment F, tests were made on trials 10 and 30 of OL and trial 1 of IL; and for Subexperiment G, tests were made on trials 10 and 50 of OL and trial 1 of IL. No *F*-ratio was significant at the .05 level.

The mean absolute decrements in matches and the mean absolute increments in errors on the first RL trial as a function of the amounts of OL and IL, are given in Table I, and are presented graphically in Fig. 1. Also included in the table are relative decrements and relative increments obtained by dividing the absolute changes by the corresponding means on the last trial of OL.

In the left-hand graph of Fig. 1, decrements in matches are plotted against amounts (days) of IL, while increments in errors are plotted against

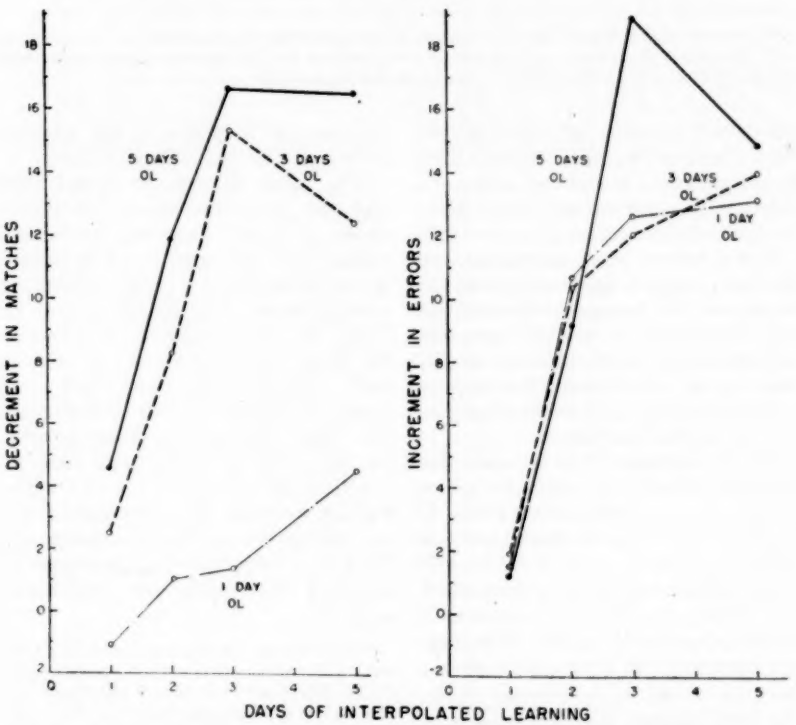

FIG. 1. Decrements in matches and increments in errors from the last OL trial to the first RL trial as a function of the days of IL following 1, 3, or 5 days of OL

TABLE I

DECREMENTS IN MATCHES AND INCREMENTS IN ERRORS FROM THE LAST TRIAL OF OL TO THE FIRST TRIAL OF RL AS A FUNCTION OF DIFFERENT AMOUNTS OF OL AND IL

Subexp.	Group	N	Days of Practice			Mean Absolute Change		Mean Relative Change**	
			OL	IL	RL	Decrement in Matches	Increment In Errors	Decrement in Matches	Increment In Errors
E	15	8	1	1	2	−1.1*	1.5	−0.053*	0.109
	30	10	1	2	2	1.1	10.7	0.052	0.930
	16	10	1	3	2	1.4	12.7	0.064	0.830
	17	10	1	5	2	4.5	13.2	0.186	0.898
F	18	10	3	1	2	2.5	1.9	0.074	0.146
	29B	10	3	2	2	8.3	10.4	0.245	1.095
	19	8	3	3	2	15.4	12.1	0.429	0.829
	20	10	3	5	2	12.4	14.0	0.354	1.069
G	21	10	5	1	2	4.6	1.3	0.108	0.092
	31A	10	5	2	2	11.8	9.2	0.296	0.780
	22	9	5	3	2	16.7	18.8	0.410	1.861
	3	11	5	5	2	16.5	14.8	0.429	1.410

* The minus sign indicates that there was an increment instead of a decrement.

** The *relative* decrements and increments were obtained by dividing the *absolute* decrements and increments by the corresponding means on the last trial of OL.

amounts of IL in the right-hand graph. Each curve in the figure depicts data for the groups in one of the subexperiments. The curves are labelled on the basis of the number of days of OL.

It will be seen that, in general, the absolute amount of interference, as represented by decrements in matches and increments in errors, increases with increases in the amount of IL. None of the inversions which appear in the curves in Fig. 1 is significant at the .10 level of confidence.

Relative decrements in matches and *relative* increments in errors are given in the last two columns of Table I. In general, the relative decrements in matches increase, for each subexperiment, with increases in the amount of IL. There is only one inversion—between Groups 19 and 20. The relative increments in errors are less consistently related to the amount of IL. The least amount of IL in each subexperiment yielded the smallest relative increment in errors, but there was at least one inversion of the relative increments in each subexperiment.

The mean decrements in matches and the mean increments in errors shown in Table I have been evaluated within each subexperiment to determine the significance of the differences among them. The results appear in Table II. The individual decrements for *Ss* in each of the four groups in each subexperiment were first subjected to simple analysis of variance[3] (**3**). The within-groups sum of squares for each subexperiment was used to calculate the standard error of the difference between the decrements for any two groups in the subexperiment. This error term was then employed in applying the *t* test for unrelated measures (**3**).

[3] The *F*-values for matches were 8.52, 12.50, and 7.87 for Subexperiments E, F, and G, respectively, with either 3 and 34 or 3 and 36 *df*. All three values are significant beyond the .001 level. The *F*-values for errors were 8.17, 6.39, and 15.82 for Subexperiments E, F, and G, and all are significant at the .01 level or better.

As will be seen, some of the differences are highly significant. This was usually the case for differences between groups having 10 trials (one day) of IL and those having 30 or 50 trials, as would be expected from the relative magnitude of the decrements in Table I and from the appearance of the curves in Fig. 1. Some of the differences between the matches decrements and between the error increments were lacking in statistical significance. However, except for the inversions already noted above, there was a general trend for the decrements and the increments to increase with increases in the amount of IL.

TABLE II

SIGNIFICANCE FOR DIFFERENCES BETWEEN THE MEANS OF DECREMENTS IN MATCHES AND THE MEANS OF INCREMENTS IN ERRORS FROM THE LAST TRIAL OF OL TO THE FIRST TRIAL OF RL

Sub-exp.	Groups Compared	Amounts of IL in Days	t****	
			Matches	Errors
E	15 & 30	1 & 2	1.93	3.49**
	15 & 16	1 & 3	2.19	4.24***
	15 & 17	1 & 5	4.91***	4.43***
	30 & 16	2 & 3	0.28	0.80
	30 & 17	2 & 5	3.18**	1.00
	16 & 17	3 & 5	2.90**	0.20
F	18 & 29B	1 & 2	2.71*	2.86**
	18 & 19	1 & 3	5.66***	3.24**
	18 & 20	1 & 5	4.63***	4.07***
	29B & 19	2 & 3	3.11**	0.54
	29B & 20	2 & 5	1.92	1.21
	19 & 20	3 & 5	1.32	0.60
G	21 & 31A	1 & 2	2.52*	2.97**
	21 & 22	1 & 3	4.12***	6.32***
	21 & 3	1 & 5	4.27***	5.19***
	31A & 22	2 & 3	1.67	3.52**
	31A & 3	2 & 5	1.68	2.15*
	22 & 3	3 & 5	0.07	1.50

* Significant at the .05 level of confidence, but less than .01.

** Significant at the .01 level, but less than .001.

*** Significant at the .001 level, or better.

**** Comparisons in Subexperiments E, F, and G based on 34, 34, and 36 *df*, respectively.

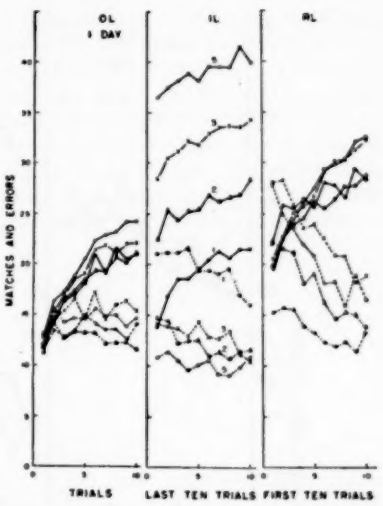

FIG. 2. Matches and errors on last 10 OL trials, last 10 IL trials, and first 10 RL trials when 1, 2, 3, or 5 days of IL follow 1 day of OL. The days of IL are indicated beside the coded curves in the center section of the figure.

The performance of groups on RL trials other than the first must, of course, be considered in formulating generalizations regarding the effect of the different degrees of interpolated learning. In Figs. 2, 3, and 4 are presented the matches and total-error curves for the first 10 RL trials after the four degrees of IL which followed 1, 3, and 5 days of OL. The left-hand graph in each figure shows the last 10 trials of OL,[4] the middle graph shows the last 10 trials of IL, and the right-hand graph shows the first 10 trials of RL.

The performance of the groups on RL trials other than the first tends to confirm the findings already mentioned. A study of the RL curves in

[4] A comprehensive and detailed record of the performance of Group 20 is presented graphically in Fig. 1 of the paper by Lewis, McAllister, and Adams (2).

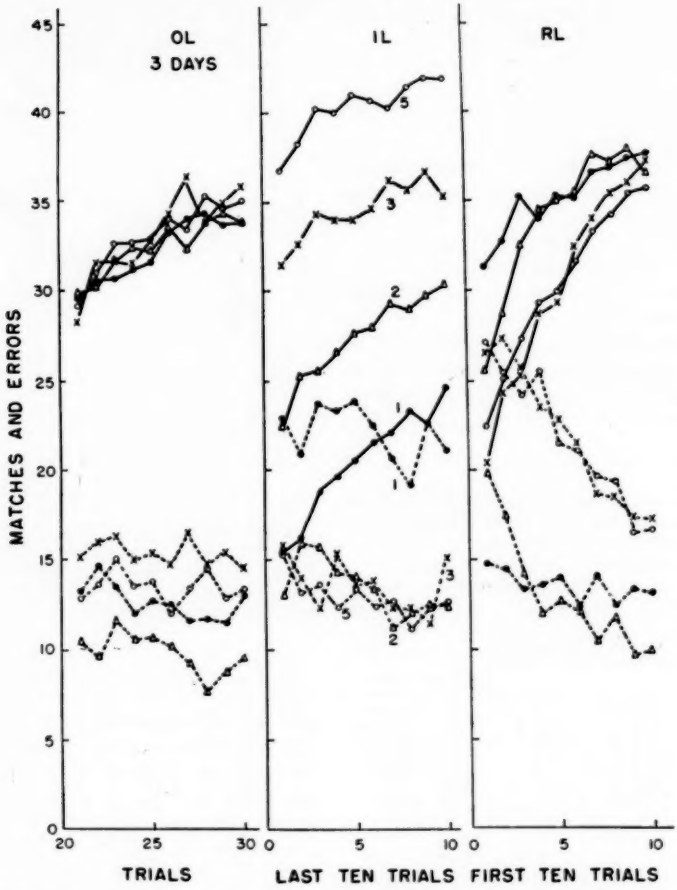

FIG. 3. Matches and errors on last 10 OL trials, last 10 IL trials, and first 10 RL trials when 1, 2, 3, or 5 days of IL follow 3 days of OL. The days of IL are indicated beside the coded curves in the center section of the figure.

Figs. 2, 3, and 4 reveals that, as a rule, the groups in each subexperiment retained the same relative positions through the first several RL trials. For some groups, the effects of the interpolated practice seemed to persist beyond the first day of RL.

DISCUSSION

Two widely known studies of interference in rote serial learning, one by McGeoch (4), the other by Melton and Irwin (5), showed that with some fixed amount of OL, an increase in the amount of IL results in an increase in the amount of retroactive inhibition, with the maximal amount of inhibition occurring with intermediate amounts of IL. The decrease in amount of inhibition at high levels of IL was not confirmed for paired-associates learning either by Thune and

Underwood (6) or by Underwood (7), but all four of the investigations revealed that the number of overt intrusions during RL decreases when IL is carried to high levels.

In the present experiment, there was some indication that the largest amount (five days) of interpolated practice induced less interference than an intermediate amount (three days), when OL had been for three or five days, but not when OL had been for only one day. However, the results must be interpreted with caution. None of the inversions of the general trend was significant. Also, the groups within the subexperiments were not brought to a uniform performance

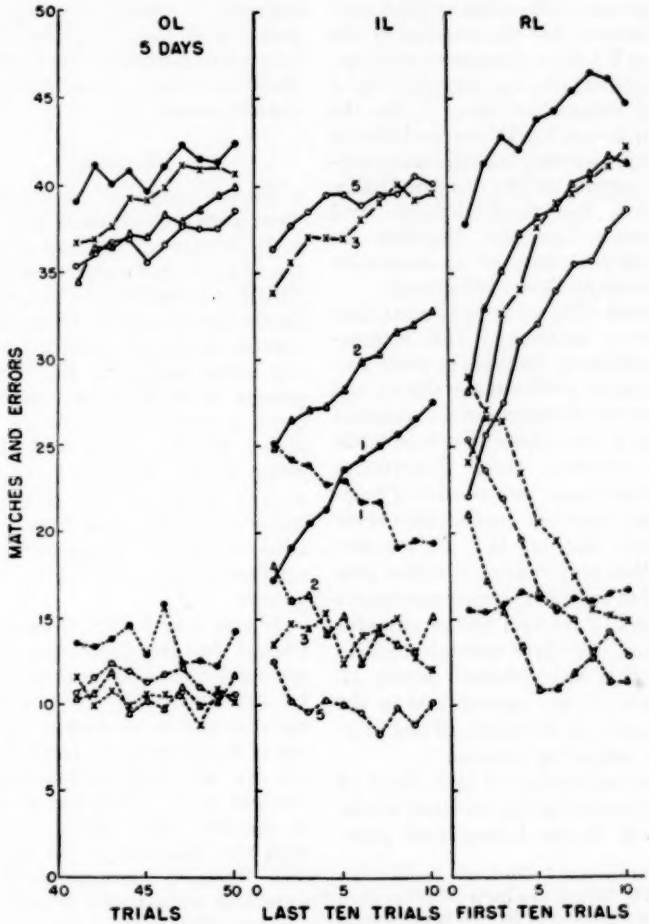

FIG. 4. Matches and errors on last 10 OL trials, last 10 IL trials, and first 10 RL trials when 1, 2, 3, or 5 days of IL follow 5 days of OL. The days of IL are indicated beside the coded curves in the center section of the figure.

level, but were given the same number of practice trials. Even though the four groups in each subexperiment were shown by analysis of variance techniques to be comparable in ability to perform, there may actually have been differences in ability which led to the inversions displayed in Fig. 1.

If later experiments on motor learning should show that, with some uniform amount of OL, some large amount of IL induces less interference at the outset of RL than some lesser amount, the results might be explained by a kind of two-factor theory.[5] In the paper by Lewis, McAllister, and Adams (2), which analyzed the effects of varying the amount of OL, it seemed likely that two factors—facilitation and interference—operated together in some complex manner to determine the over-all level of performance.

It seems reasonable to suggest that retroactive facilitation and retroactive interference combine in some way to determine performance during the initial stages of relearning the standard Mashburn task subsequent to practice on the reversed task. Retroactive interference may well consist of inappropriate response tendencies—those developed during IL. Retroactive facilitation may consist of either generalized or specific response tendencies developed in relation to the stimulus lights and the three controls tendencies further strengthened during IL which are entirely appropriate to the performance of the standard task during the relearning process.

In the acquisition of high levels of skill in performing the reversed Mashburn task during interpolated practice, Ss may very conceivably develop tendencies to respond which contribute positively to the performance of the standard task. If they do, retroactive facilitation might be expected to become apparent (that is, to override some of the effects of retroactive interference) only after relatively large amounts of IL.

The foregoing discussion should be regarded as highly tentative. A future investigation will be aimed at determining whether the inversions in the present data are genuine or mere chance variations.

[5] Not to be confused with the two-factor theory developed by Melton and Irwin (5) primarily to account for the observed decrease in the amount of retroactive inhibition which accompanied increases in the amount of IL beyond an intermediate amount.

Summary

Twelve groups of Ss, divided into three subexperiments of four groups each, were given different amounts of OL and different amounts of IL on the Modified Mashburn Apparatus, to determine the effects of variations in amount of IL on performance during the initial stages of RL. The four groups in each subexperiment were given a fixed number of OL trials—either 10, 30, or 50—and then had either 10, 20, 30, or 50 IL trials. All groups relearned the original task under the same conditions. The results could be interpreted without reference to the performance of control groups.

It was found that, in general, for each of the three fixed amounts of OL, retroactive interference, as measured by decrements in mean number of matches and increments in mean number of errors from the last OL trial to the first RL trial, increased with the amount of IL. Relative decrements in matches also tended to increase with the increases in the amount of IL. Relative increments in errors were less consistently related to the amount of IL.

There were two sizeable inversions in the trends of the mean decrements

in matches and the mean increments in errors. These inversions suggest that relatively large amounts of IL may tend to induce smaller decrements in matches and smaller increments in errors than some lesser amount of IL. However, the inversions were not statistically significant. Furthermore, a firm conclusion on this point could not be reached because Ss within each subexperiment were not brought to a uniform performance level during OL but were given an equal number of practice trials, making it impossible to state with assurance that the groups were equal in ability to perform.

(Manuscript received March 27, 1950)

References

1. Lewis D., and Shephard, A. H. Devices for studying associative interference in psychomotor performance: I. The modified Mashburn apparatus. *J. Psychol.*, 1950, **29**, 35–46.
2. Lewis, D., McAllister, Dorothy E., and Adams, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
3. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
4. McGeoch, J. A. The influence of degree of interpolated learning upon retroactive inhibition. *Amer. J. Psychol.*, 1932, **44**, 695–708.
5. Melton, A. W., and Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 175–203.
6. Thune, L. E., and Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, **32**, 185–201.
7. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**, 33 pp.

TWO-POINT DISCRIMINATION IN VISUAL SPACE AS A FUNCTION OF THE TEMPORAL INTERVAL BETWEEN THE STIMULI [1]

BY MICHAEL LEYZOREK [2]

The Johns Hopkins University [3]

In classical studies of spatial discrimination the stimuli, be they lines, points, or spaces, are seen simultaneously by the observer while the judgment is made. The problem dealt with here concerns the precision of localization when the points to be discriminated are separated in time as well as in space. This problem can be elaborated as follows. Let S_1 and S_2 be two easily visible points of short duration which are presented successively in different positions. When the interval, T, between S_1 and S_2 equals zero, the two points are seen simultaneously and discretely if the distance, D, between them is discriminable. In this case a simple spatial discrimination is made. As T increases from zero to approximately .02 sec., the perceptual situation remains unchanged, but at .02 sec. apparent movement of S_1 begins to appear if D is small enough. For values of T between .03 and .20 sec. the apparent movement of S_1 to S_2 is very evident. As T increases beyond .20 sec., S_1 and S_2 are perceived discretely in time without apparent movement, and the discrimination of their relative positions becomes the problem under consideration here.

A number of factors may affect this kind of judgment. Differences in the characteristics of the retina from one point to another, eye movements between the presentation of S_1 and S_2, and the number of cues or references available in the visual field are some of these. The nature of the memory process must also be considered. Other things equal, memory for the position of a point should deteriorate in the same general fashion as is classically predicted for any memory trace, namely, as a negatively accelerated exponential function. Within reasonable limits, characteristics such as size, brightness, color, and shape of the stimuli and background should be of secondary importance.

Of the many relevant parameters only two were selected for study of their effect on this function. These were: (*a*) two different degrees of visual field complexity and (*b*) foveal observation with uncontrolled eye fixation versus peripheral observation with controlled eye fixation.

Earlier work on this problem.—A search of the literature has unearthed no prior attack on the problem of spatial discrimination as it is here formulated. Although many studies are peripherally related to this problem, only one is of direct interest. Nixon (2), investigating immediate memory for spatial position, presented *S*s with patterns of black dots on a white cir-

[1] This investigation was performed under Contract N5-ori-166, Task Order I, between the Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-125, Project Designation No. NR-784-001, under that contract.

[2] Now at Psychology Branch, Naval Research Laboratory, Washington 20, D.C.

[3] The writer is indebted to Drs. A. Chapanis, W. R. Garner, and J. W. Gebhard for their helpful suggestions concerning this problem. Mr. E. J. King gave invaluable advice and assistance in the design and construction of the apparatus. Many thanks are due the *S*s for their diligent and careful efforts in an arduous task.

cular field. After intervals of 3, 5, and 10 min., the *S*s attempted to reproduce the positions of the dots on sheets of paper. Nixon found that errors of reproduction increased with time before recall and with the number of dots in the pattern being recalled. Individual differences between *S*s and the position of the dots with respect to the frame of reference exerted less influence on the accuracy of recall.

Apparatus, Subjects, and Procedure

The visual problem.—The visual problem presented to the *S*s was a square white screen 4 ft. on a side, seen at the end of a long, narrow room as shown in Fig. 1. The viewing distance was 12 ft. The points, S_1 and S_2, appeared in the central area of the screen as small white dots, each subtending a visual angle of slightly less than 3 min. of arc. The *S* viewed the screen binocularly and without restriction of head movement. This manner of observation was specified in order to get the desired data under "normal" or ordinary conditions of visual search. Occasionally it was found that *S* obtained head support by use of his forearms and hands.

Depending upon his instructions, *S* observed the stimuli either foveally or peripherally with the help of a fixation light. When the fixation light was used, it appeared continuously as a small red spot of light to the right of the center of the screen. In this situation *S* saw S_1 and S_2 approximately 5° to the left of the fixation light.

Using either foveal or peripheral observation as required, *S* made judgments both in light and in complete darkness. When judgments were made in light, the screen was diffusely front-lighted from sources invisible to *S*. The screen surface itself was perfectly homogeneous at the 12-ft. viewing distance, but the screen boundaries and details of the interior of the experimental room were clearly visible. Because of the front-lighting, nothing behind the screen could give cues to *S* as to the probable location of the stimulus points.

No visual cues were available to *S* when he worked in complete darkness. In this situation the screen itself could not be located in advance of the presentation of S_1 and S_2. With the fixation light *S* could tell approximately where the stimuli would appear but still could not see the screen.

To allow *S* sufficient time for finding S_1 and fixating it under all conditions of observation, S_1 was given a duration of slightly more than 1 sec.; S_2 had a duration of 0.5 sec.

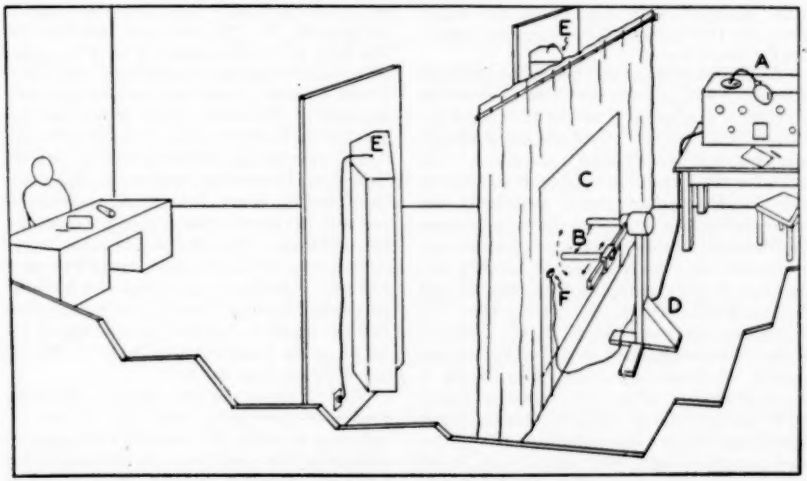

Fig. 1. A view through the side of the experiment room showing details of the apparatus and its arrangement with respect to the subject

That part of the room behind the screen (occupied by apparatus and *E*) was completely masked off from *S* by black curtains extending from the edges of the screen to walls, floor, and ceiling.

Conditions of observation.—In summary, the four experimental conditions of observation were: (A) Stimuli presented on a brightly illuminated screen, foveal observation, no control over fixation; (B) Stimuli presented on an illuminated screen, peripheral observation, controlled fixation; (C) Stimuli presented in complete darkness, foveal observation, no control over fixation; and (D) Stimuli presented in complete darkness, peripheral observation, controlled fixation.

Apparatus.—The visual display presented to *S* was accomplished with five major items of apparatus shown in Fig. 1. (*a*) An electronic light-flash generator of special design (**3**). This provided for precise control of the duration of S_1 and S_2 and of the interval, *T*, between them. (*b*) Two flash tubes (Sylvania Model R 1131A), fired by the flash generator.

These tubes were selected for their sharp on-off characteristic, this being almost perfectly square. (*c*) A 4-ft. square translucent white screen on which the output from the flash tubes was projected by a suitable lens system. This glass screen was mounted so that it could be raised or lowered to bring its central point to the approximate eye height of the *S* viewing it. (*d*) A mechanical system for positioning the flash tubes and the fixation light behind the screen. (*e*) Two vertically mounted 40-w. fluorescent lamps (GE "daylight") 40 in. long, one on each side of the screen frame, provided the diffuse front-lighting for the screen.

Additional contrast was needed between the stimulus points and the screen background for Condition B in which *S* viewed the stimuli peripherally in the light. To obtain this condition, the amount of front-lighting was reduced. The need for the increase in brightness contrast in this situation was determined empirically, but experimental data concerning the sharp decrease in differential sensitivity in the periphery as compared with the fovea have been reported by several investigators, among them Wallace (**4**), and Lamar, Hecht, Schlaer, and Hendley (**1**). The luminance values at which S_1, S_2, and the background were maintained in the various experimental conditions are presented in Table I. These luminance values were established empirically and gave stimuli which were always clearly visible but did not leave undesirable after-effects such as after-images, blinding glare, or, in the two experimental conditions in which *S* worked in the dark, loss of dark adaptation.

Procedure.—The four experimental conditions described above were presented to *S* in counterbalanced order. Within each experimental condition a separate method-of-constants experiment was done with each *S* to obtain a spatial discrimination threshold for each of the six temporal intervals studied. The temporal intervals chosen ranged from .25 to 16.0 sec. in six equal logarithmic steps (ratio = 2.297) as follows: .25, .57, 1.32, 3.03, 6.96, and 16.0 sec. The lower limit of this range (T = .25 sec.) was selected as the shortest temporal interval between S_1 and S_2 which would not produce apparent movement. The choice of the upper limit was based on the results of several preliminary experiments. In one of these, six intervals in equal log steps from .25 to 8.0 sec. were tested, and it was found that the curve of the resulting thresholds was very flat between the two longest intervals, 4.0 and 8.0 sec. To explore this function further in time, it was decided to add a longer interval of 16.0 sec. Additional experiments with the range thus extended showed that six points from .25 to 16.0 sec. would be sufficient to demonstrate the nature of the function under study—hence the use of the six steps indicated.

Table II shows the order in which the major conditions were tested with each *S* and the sequences in which the temporal intervals were assigned within conditions. As indicated in this table, the *S*s worked their way through each of the constants series until they had completed the six series required within a major condition, e.g.,

TABLE I

LUMINANCE VALUES FOR STIMULUS POINTS AND SCREEN

Experimental Condition	Luminance in Apparent Foot-Candles	
	Stimulus Points	Screen
A. Foveal observation in the light, uncontrolled fixation	505.00	220.0
B. Peripheral observation in the light, controlled fixation	293.00	7.8
C. Foveal observation in the dark, uncontrolled fixation	2.85	0.0
D. Peripheral observation in the dark, controlled fixation	2.85	0.0

Condition D. Then they repeated the process for the next experimental condition, e.g., C, and so on until six complete sets of method-of-constants data were obtained for each S in each of the four experimental conditions.

Prior to each method-of-constants study for a given temporal interval, a preliminary method of limits was used to determine the appropriate range of distance, D, between S_1 and S_2 which would bracket the S's spatial discrimination threshold for the given value of T. Seven equidistant points in this range were used as the final stimulus values. Twenty judgments were made for each value of D. Thus, each final threshold value obtained in a constants series was based on 140 judgments.

In an attempt to increase the precision of the data, S was required to determine the relative position of S_2 with respect to S_1. S_1 appeared at the approximate center of the viewing screen, and S_2 was presented in one of four positions with respect to S_1, either above, below, to the right, or to the left. S_2 was presented in each of these orientations an equal number of times, the directions being randomized throughout a single experiment or constants series. The S's judgments were forced, since he was never allowed to report that the stimuli were in the same place.

Subjects.—Four Ss participated in this study. These were young adults, three female and one male, from 18 to 30 years old. The vision of all Ss was found to be within normal limits.

Treatment of the data.—Within each constants series (e.g., Condition A, $T = .25$ sec., D presented in seven equal steps from 1 to 7 min. of visual angle), the judgments of S were scored, and the proportion of correct judgments was corrected for chance success by the use of the formula $P_a = \frac{4}{3}(C/N) - \frac{1}{3}$, where P_a = adjusted proportion of correct responses, C = number of correct judgments, and $N = 20$. It will be noted that this correction formula assumes no bias in the Ss' judgments with respect to the four directions of presentation of S_2. Upon examination of the data for indications of such bias, it was found that the Ss did not distribute their judgments equally often among the four directions of stimulus presentation. This matter is dealt with more fully later. Suffice it to say here that the Ss' biases in this respect did not appear to distort the data in any noticeable degree, and consequently they were not taken into account in the correction formula.

The adjusted proportions were assumed to be points on the probability integral of a normal frequency distribution and, for a given set of constants data, were converted to z scores. The z scores were used as y values in a conventional unweighted least-squares solution for the best-fitting line. The median value of D in each such function, obtained by solving for D at the point where $z = 0$, was the final threshold value for that particular constants series. Ninety-six such threshold values were calculated.

TABLE II

Experimental Design

The Arabic numbers in the cells below specify the order in which the major experimental conditions were presented to each observer. Each Roman number stands for one of the six temporal intervals tested,* and they are arranged from left to right in the sequence in which the corresponding temporal intervals were presented within a major condition. See text for a description of the experimental conditions.

Observer	Experimental Condition			
	A	B	C	D
B. B.	1 V II IV I III VI	2 III V II VI I IV	4 VI I III V II IV	3 II IV VI I V III
H. B.	4 IV I VI II V III	1 I V III VI IV II	3 IV VI I III II V	2 III VI V IV II I
J. B.	3 IV II V III I VI	4 V II III I VI IV	2 III V VI I IV II	1 II V IV I III VI
R. S.	2 IV VI I III II V	3 II IV VI I V III	1 I III V IV VI II	4 V II IV I III VI

* I = 0.25 sec.; II = 0.57 sec.; III = 1.32 sec.; IV = 3.03 sec.; V = 6.96 sec.; VI = 16.00 sec.

Results and Discussion

The complete mass of data is shown in Table III, which also lists the mean threshold values for the four *Ss* for all the values of T within each condition. These mean threshold values are plotted in Fig. 2.

From the data of each least-squares solution a measure was obtained of the standard deviation of each *S*'s distribution of judgments around the mean or threshold value of D. This variability measure, σ, is an index to the range of D values necessary, within a given temporal interval, to get a normal proportion distribution of correct judgments from approximately zero per cent correct at one end to approximately 100 per cent correct at the other. These data are presented in Table IV, which includes the mean σ values for all *Ss* and conditions for each value of T. These mean σ values are plotted in Fig. 3.

Accuracy of discrimination.—The first item of interest in the data is the extraordinary precision achieved by the *Ss* in their localization of S_2. For example, as indicated in Fig. 2, the *Ss* working in Condition A (foveal observation in the light, uncontrolled fixation) showed that they can discriminate the positions of S_2 with respect to S_1 when $D = 2.23$ min. of visual angle and $T = .25$ sec. The threshold increases slowly as T increases and levels off at $D = 11.74$ min. of visual angle between $T = 6.96$ and $T = 16.0$ sec. Thus, even when

TABLE III

Threshold Data in Minutes of Visual Angle for All Experimental Conditions and Temporal Intervals

Conditions	Temporal Interval (sec.)	Observers				
		B. B.	H. B.	J. B.	R. S.	Mean
A. Foveal observation in the light, uncontrolled fixation	0.25	2.62	1.65	3.23	1.41	2.23
	0.57	5.17	1.70	4.72	1.77	3.34
	1.32	4.58	2.25	5.00	3.49	3.83
	3.03	7.97	5.90	9.49	5.64	7.25
	6.96	10.84	9.86	14.46	11.80	11.74
	16.00	12.93	10.87	14.16	8.50	11.62
B. Peripheral observation in the light, controlled fixation	0.25	8.53	7.70	7.67	5.63	7.38
	0.57	10.54	7.88	8.35	9.66	9.11
	1.32	13.47	11.53	8.45	10.00	10.86
	3.03	10.00	7.34	5.56	8.75	7.91
	6.96	11.12	8.78	14.47	8.31	10.67
	16.00	11.83	8.79	9.42	10.64	10.17
C. Foveal observation in the dark, uncontrolled fixation	0.25	7.83	7.01	8.79	6.16	7.45
	0.57	11.19	7.62	9.07	8.91	9.20
	1.32	27.53	30.37	24.34	27.19	27.36
	3.03	42.41	51.37	71.97	54.68	55.11
	6.96	105.57	78.15	182.24	106.89	118.21
	16.00	168.48	138.03	241.34	186.84	183.67
D. Peripheral observation in the dark, controlled fixation	0.25	9.31	6.69	9.03	5.69	7.68
	0.57	14.57	9.99	19.49	5.09	12.29
	1.32	11.45	6.23	6.48	10.82	8.75
	3.03	13.49	11.05	11.21	11.04	11.70
	6.96	13.10	8.28	11.84	11.27	11.12
	16.00	17.55	9.94	20.41	8.77	14.17

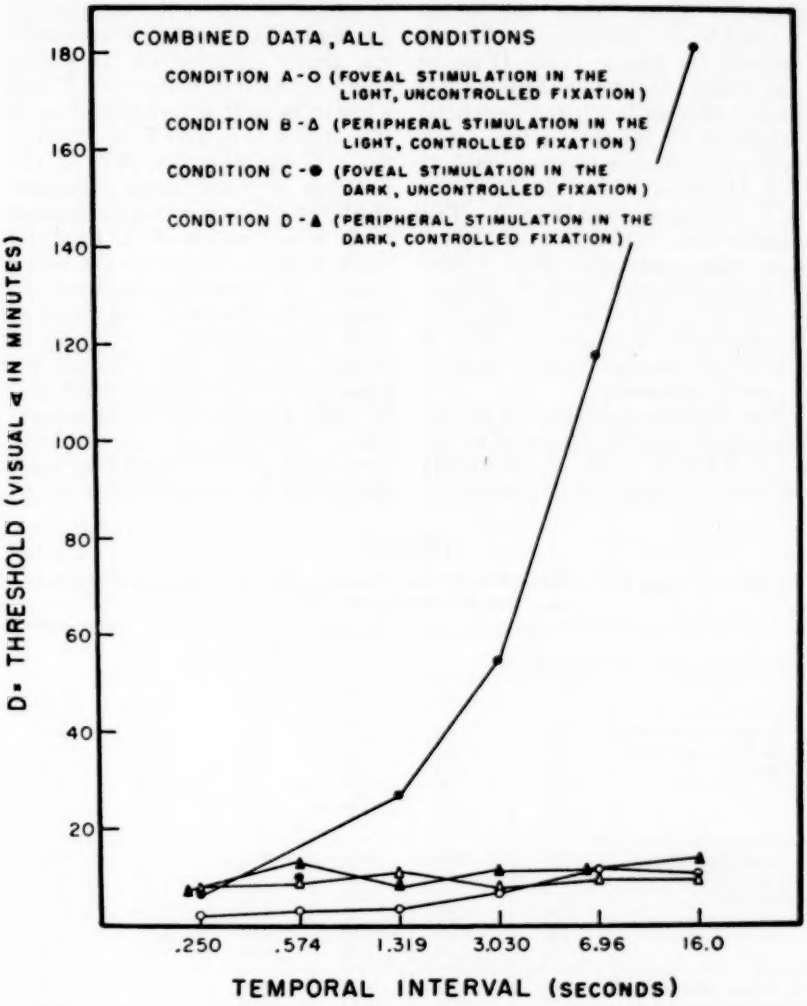

FIG. 2. Mean spatial discrimination thresholds in minutes of visual angle for all *S*s, plotted as a function of the temporal interval between the stimuli

16 sec. elapses between the presentation of two stimulus points, *S* can discriminate their relative positions with accuracy comparable to that achieved by many individuals in ordinary tests of visual acuity.

The thresholds obtained with peripheral observation in Conditions B (peripheral observation in the light, controlled fixation) and D (peripheral observation in the dark, controlled fixation) might be expected to be con-

siderably higher than the foveal thresholds of Condition A. Examination of the data in Table III and of Fig. 2 shows that while there is a difference between foveal and peripheral thresholds for the shorter time intervals, this difference becomes small at $T = 3.03$ sec. and virtually disappears at $T = 6.96$ sec. and beyond. This suggests that, in this kind of observation, when peripheral vision is supported by a fixation point, the characteristic inferiority of parafoveal spatial discrimination to that of the fovea is rapidly overcome as the delay between S_1 and S_2 increases.

The probable importance of the fixation point is further suggested by the quite different order of thresholds obtained in Condition C (foveal observation in the dark, uncontrolled fixation). In this case the thresholds rise almost exponentially as a positively accelerated function of T from a low value of $D = 7.45$ min. of arc at $T = 0.25$ sec., to over 3° at $T = 16.0$ sec. (If the rising threshold in this situation were considered a measure of S's "pure" memory for the position of S_1 as a function of T, then this curve could be considered the mirror image of a decreasing, negatively accelerated function describing the decay of a memory trace.) It should be pointed out here that a spatial discrimination threshold of as much as 3° of visual angle is still rather remarkable in this situation. In Condition C not only were no visual cues available to the Ss to help localize S_1 and

TABLE IV

VARIABILITY DATA (σ) IN MINUTES OF VISUAL ANGLE FOR ALL EXPERIMENTAL CONDITIONS AND TEMPORAL INTERVALS

Conditions	Temporal Interval (sec.)	Observers				
		B. B.	H. B.	J. B.	R. S.	Mean
A. Foveal observation in the light, uncontrolled fixation	0.25	1.88	1.23	2.05	0.61	1.44
	0.57	3.12	1.36	2.03	0.70	1.80
	1.32	2.62	3.56	2.69	1.63	2.62
	3.03	5.27	4.39	4.42	4.69	4.69
	6.96	6.55	7.54	7.05	5.29	6.61
	16.00	5.22	6.21	10.65	6.91	7.25
B. Peripheral observation in the light, controlled fixation	0.25	5.03	4.95	6.22	5.70	5.48
	0.57	4.73	5.52	7.09	6.73	6.02
	1.32	9.62	5.41	5.86	6.11	6.75
	3.03	6.15	7.82	7.15	7.45	7.14
	6.96	9.28	4.98	7.84	4.29	6.60
	16.00	7.59	8.75	7.66	9.08	8.27
C. Foveal observation in the dark, uncontrolled fixation	0.25	4.62	4.44	4.06	3.95	4.27
	0.57	2.76	4.42	5.19	3.85	4.06
	1.32	22.88	20.90	20.09	18.66	20.63
	3.03	34.35	54.38	45.04	30.22	41.00
	6.96	65.79	44.82	96.15	67.25	68.50
	16.00	101.21	99.30	93.20	95.15	97.22
D. Peripheral observation in the dark, controlled fixation	0.25	4.75	4.76	6.14	6.35	5.50
	0.57	6.06	10.15	6.68	10.15	8.26
	1.32	6.45	4.86	4.29	5.76	5.34
	3.03	9.68	5.20	6.66	4.53	6.52
	6.96	8.64	7.02	6.76	5.80	7.06
	16.00	8.26	5.65	14.35	5.80	8.52

S_2, but apparently the remaining aids to orientation, e.g., generalized kinesthetic and proprioceptive stimuli, were insufficient for locating the screen itself. On a few occasions *Ss* missed S_1 completely because they were looking in the wrong direction, thus requiring a repetition of the stimuli.

Visual field complexity.—What of the role played by the degree of complexity or structuring of the visual field? For purposes of discussion let "complexity" here be defined in terms of the number and arrangement of clearly resolvable inhomogeneities in the field of vision. Granting the difficulty of defining as well as specifying quantitatively the amount of "complexity" or "structuring" in a visual field, we do not think it unreasonable to rank the visual stimulus pattern of Condition A (foveal observation in the light, uncontrolled fixation) as more structured than that of Condition C (foveal observation in darkness, uncontrolled fixation). Obviously, the visual field pattern presented in Condition A, while more complex than that of Condition C, is still relatively simple by comparison with most visual stimulus patterns encountered in ordinary circumstances. A comparison of the thresholds obtained in these two conditions of observation clearly indicates the expected superiority of judgments made with the help of a visible and clearly framed background.

However, another comparison involving "degree of complexity" gives different results, i.e., between Conditions B (peripheral observation in the light, controlled fixation) and D (peripheral observation in the dark, controlled fixation). In Condition B the visual stimulus pattern is essentially that in Condition A with the addition of a fixation point, while Condition D resembles Condition C, with a fixation point added. Thus, Condition B is presumably more complex than D, but the absolute threshold values in these two functions are very similar. In this case, the structuring produced by the fixation light alone (Condition D) appears to be as powerful a factor in maintaining good spatial discrimination as the presence of a more complex visual field (Condition B), produced by general illumination of the experimental room.

These apparently conflicting findings concerning the relationship of visual field complexity to the accuracy of localization suggest a need for more precise specification of this factor. It is not unreasonable to suppose that while spatial discrimination as practised here may be a function of visual field complexity, such a function is not likely to be monotonic. That is, there may be some optimum degree of visual field complexity for spatial localization with inferior discrimination resulting from simpler or more complex structuring.

Variability.—By comparing the variability data of Table IV and Fig. 3 with the threshold data of Table III and Fig. 2, it can be seen that the *Ss*' judgments are more variable as they decrease in precision. The r between the two measures is high, and the proportionality between them is nearly constant with a ratio of roughly 1:1.5 through the entire range of values of T.

Position bias.—It was frequently reported by the *Ss* that S_2 did not appear directly on the rectangular coordinate specified for it, namely, up, right, down, or left with respect to S_1. If these positions are given angular designations of 0°, 90°, 180°, and 270°, then S_2 was very commonly seen in an intermediate angular position such as 45° when S_2 was actually at 0°. These reports, together with *E*'s observation that, when dealing with difficult or

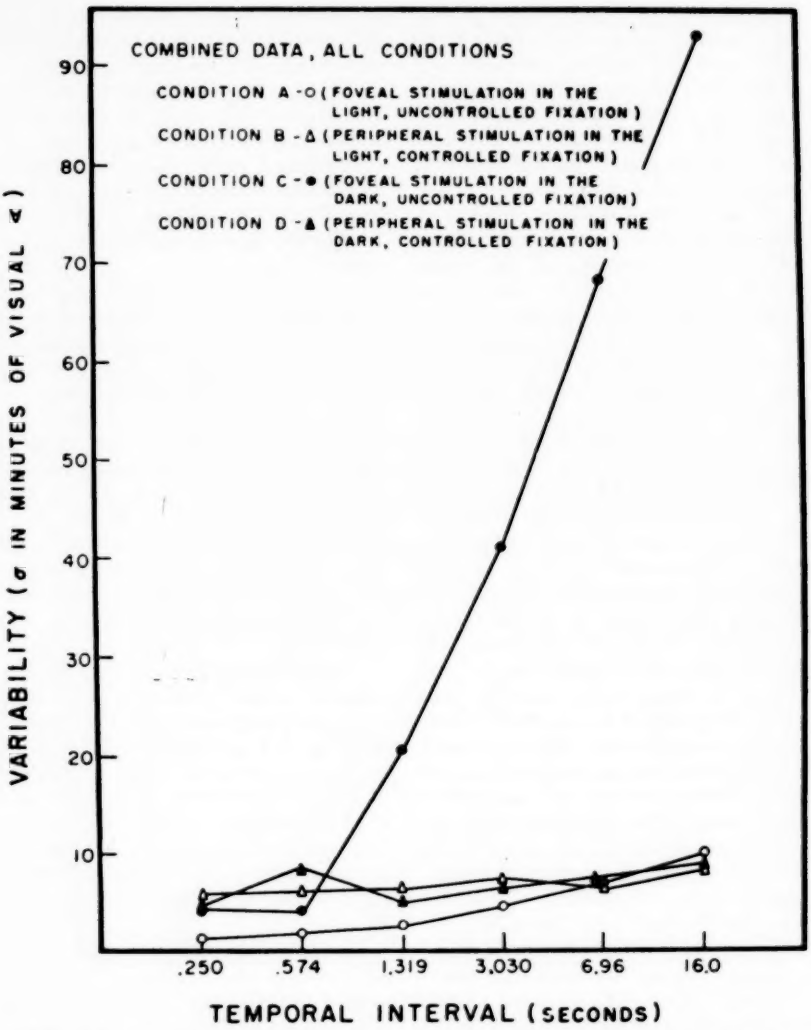

FIG. 3. Mean standard deviations in minutes of visual angle around the *Ss*' threshold values (all *Ss*) plotted as a function of the temporal interval between the stimuli

doubtful judgments, some *Ss* consistently reported S_2 in one direction to the exclusion of the others, led to an analysis of the data for some kind of position bias. This was done by tallying the number of times S_2 was reported in each of the four directions (U, D, R, L) against the number of times it was actually presented in each direction. These data are presented in Table V. If there is no tendency or set to see S_2 in one particular direc-

tion, the ratio of directions judged to directions presented should be approximately the same for all four directions of presentation. Two of the *S*s (B. B. and J. B.) did show a distinct tendency to judge S_2 as being to the right of S_1 or above it more often than in any other position.

If the data are massed for all *S*s and distributed by major experimental conditions, a slight preponderance of judgments of U and R is again seen, but this is probably due to the bias found in the data of B. B. and J. B. Condition B appeared to be most free of such tendency. Since Conditions A and C imposed no restriction on *S*'s eye movements, these should have been most favorable to the appearance of biased judgments of the direction of presentation of S_2. When the data are recombined so as to compare the judgments made in Conditions A and C with those of B and D, a preponderance of R and U judgments is found in the data of Conditions A and C, suggesting that these were, in fact, conditions favorable to internal position bias in the *S*s. A similar comparison in which the data of A and B combined are contrasted with those of C and D combined shows a larger bias

TABLE V

FREQUENCY OF THE *S*s' JUDGMENTS OF THE DIRECTION OF S_2 WITH RESPECT TO S_1, COMPARED WITH THE ACTUAL FREQUENCY OF PRESENTATION IN EACH DIRECTION
(U = Up, D = Down, R = Right, and L = Left)

Collected data for individual observers, data of all experimental conditions included						Collected data for different experimental conditions, data of all observers included			
		Observer B. B.				*Condition A*			
		REPORTED				REPORTED			
		U	D	R	L	U	D	R	L
PRESENTED	U	536	29	186	89	583	23	142	92
	D	32	515	203	90	51	495	241	53
	R	81	102	601	56	88	69	658	25
	L	114	113	128	485	117	114	77	532
	Total	763	759	1118	720	839	701	1118	702
		Observer H. B.				*Condition B*			
		U	D	R	L	U	D	R	L
PRESENTED	U	577	30	117	116	591	9	109	131
	D	51	556	123	110	18	576	125	121
	R	96	120	575	49	110	103	532	95
	L	130	103	50	557	77	141	72	550
	Total	854	809	865	832	796	829	838	897
		Observer J. B.				*Condition C*			
		U	D	R	L	U	D	R	L
PRESENTED	U	658	9	112	61	609	38	115	78
	D	42	476	263	59	67	490	180	103
	R	131	42	624	43	137	92	553	58
	L	165	97	109	469	200	87	71	482
	Total	996	624	1108	632	1013	707	919	721
		Observer R. S.				*Condition D*			
		U	D	R	L	U	D	R	L
PRESENTED	U	599	28	87	126	587	26	136	91
	D	34	554	134	118	23	540	177	100
	R	135	96	515	94	108	96	572	64
	L	86	132	68	554	101	103	135	501
	Total	854	810	804	892	819	765	1020	756

(in favor of R and U judgments) in Conditions C and D than in Conditions A and B. This suggests that the presence in Conditions A and B of a more complexly structured visual field acted to reduce the effect of the internal bias of *S* in those situations.

Another interesting aspect of the data is found on examining the distribution of incorrect judgments for any single direction of presentation. For example, using the coordinate system mentioned earlier, it can be seen that for any given stimulus, the response 180° opposite to it was very rarely given in comparison with those at 90°. This is a reasonable finding if we consider the *S*'s perceptual problem. In those instances in which *D* was small and the judgment of the position of S_2 was difficult, *S* may well have been able to see S_2 as discrete from S_1 without being certain of the direction of displacement. In such difficult stimulus situations *S* appeared to be able to assign S_2 to within 90° of its true position. This suggests that the thresholds obtained in this situation are probably somewhat higher than they might be if secured from *S*s under instructions to report the classical criteria of "oneness" or "twoness" concerning the relative position of S_1 and S_2.

Head and eye movements of Ss.—In making their judgments the *S*s were instructed to fixate the position of S_1 during the interval preceding S_2, but reports obtained from the *S*s indicate that this procedure was not followed consistently. Regardless of the method of observation specified and particularly during the longer intervals, the *S*s apparently found it difficult to maintain steady fixation on the point at which S_1 appeared. To relieve tension during the delay period, they frequently either closed their eyes or turned their heads away from the screen for some part of the time before the presentation of S_2. In such instances their judgment of the position of S_2 was apparently made with respect to a memory trace of the position of S_1 and/or of the entire visual field pattern, rather than in terms of a continuing stimulus representing the position of S_1. In view of this unplanned irregularity in the mode of observation of judgment, the precision and consistency of the *S*s, judgments are again noteworthy.

Relationship of this kind of judgment to other tasks.—Judgments of the kind studied here are common to many tasks involving a visual search for orientation points in space. For example, the seaman who observes the position of a revolving or intermittently flashing beacon in order to check on changes in his own position must be able to localize each successive flash of light accurately. His ability to do this depends, among other things, on the repetition rate of the light flashes, or on the temporal interval between them. Similarly, the radar operator who attempts to identify or track a target represented by a weak signal seen momentarily on a radar screen must compare the position of each signal or pip with his memory of the position of preceding pips, as revealed by successive rotations of the scanning beam or sweep. It is only when the location of each pip can be fixed by a continuously visible marker of some kind that the discrimination of the relative position of successive pips is independent of the time elapsing between them.

For spatial orientation or discrimination problems in which the visual reference points are moving and are not continuously available to the *S*, the data of this experiment might be used to specify the maximum allowable delay between "looks" at the ref-

erence points for a given level of discrimination accuracy. On the other hand, given the frequency of observation of the spatial references, these data make it possible to predict the accuracy of localization.

Summary and Conclusions

1. Four *Ss* were asked to judge the position of the second of a pair of light flashes with respect to the first when the distance and the temporal interval between these stimulus points were varied systematically. The temporal intervals were .25, .57, 1.32, 3.03, 6.96, and 16.0 sec. Four conditions of observation were used:

A. Foveal observation of stimulus points presented on a brightly lighted screen with no control of the *Ss*' fixations.

B. Peripheral observation of stimuli presented on a lighted screen with the *Ss*' eye movements controlled by a fixation point 5 deg. to the right of the stimulus area.

C. Foveal observation of stimulus points presented in complete darkness, fixation not controlled.

D. Peripheral observation in complete darkness, fixation controlled as in (B).

2. The range of thresholds obtained, from the shortest time interval, .25 sec., to the longest, 16.0 sec., was as follows:

a) Condition A gave the lowest thresholds, 2.2 to 11.6 min. of arc.

b) Conditions B and D gave approximately the same thresholds, from 7.4 to 14.2 min. of arc.

c) Condition C gave the highest thresholds, from 7.5 min. to 3 deg. 4 min. of arc.

d) The variability of the *Ss*' judgments was inversely proportional to their precision.

3. It is concluded that:

a) Two points can be discriminated in visual space with high precision even when the points are presented successively, separated by intervals of as long as 16 sec.

b) Discrimination is better when the points are presented on a clearly framed background than when they occur without visual reference points, as in the dark.

c) A single fixation point is as effective as a more complex visual pattern in maintaining the precision of judgments. For the longest temporal interval (T = 16.0 sec.) the fixation point eliminates the difference between foveal and parafoveal discrimination thresholds.

(Manuscript received for immediate publication February 15, 1951)

References

1. Lamar, S., Hecht, S., Schlaer, S., and Hendley, H. F. Size, shape, and contrast in the daytime detection of targets. *J. opt. Soc. Amer.*, 1946, **36**, 714.

2. Nixon, S. R. Some experiments on immediate memory. Medical Research Council, Applied Psychology Research Unit, Psychological Laboratory, Cambridge, England. Report No. A.P.U. 39, November, 1938.

3. Roush, R. G., and Hamburger, F., Jr. A light flash generator. *Electronics*, 1948, **21**, 100–102.

4. Wallace, S. R. Intensity discrimination in the peripheral retina. *Psychol. Bull.*, 1940, **37**, 552–553. (Abstract)

THE INFLUENCE OF FIGURE–GROUND RELATIONSHIPS IN BINOCULAR RIVALRY [1]

BY LAWRENCE T. ALEXANDER

The Johns Hopkins University [2]

When corresponding areas of the two retinas are stimulated simultaneously by different colors or by figures which differ in some respects, then one of three things may occur: the fields may fuse; only one may be seen to the exclusion of the other; or the fields may alternate. Thus, if slightly different plane projections of a solid object are presented stereoscopically to corresponding areas of each eye, the two images will fuse and produce the perception of depth. But if one of the fields contains a figure of some sort while the other is homogeneous, the figure will be seen continuously. If these two cases represent the extremes of a continuum, between them are conditions in which the fields alternate. This alternation has been called binocular rivalry.

Rivalry can be observed in two forms: "successive rivalry," in which case the alternation is complete and the two fields supercede each other completely, and "simultaneous rivalry," where both fields are seen at the same time but present a patchy appearance, i.e., only one or the other of the fields is visible at any one place in the visual field. In normal observation of stereoscopic rivalry either or both of these types may occur.

Early investigators sought to explain the phenomenon of binocular rivalry either in terms of gross eye movements or in terms of some local or central inhibitory state. Breese (**1**) reported that all efforts on the part of his *Ss* to influence rivalry were accompanied by eye movements (in the same direction of the movement of regard), but that when eye movements were prevented, rivalry continued, although all sense of "will" disappeared. Peckham (**6**) could find no relationship between eye movements and rate of alternation of phase length of rivalling fields.

Breese (**1**) has shown that the rate of fluctuation of rivalling fields varies directly as the brightness of the incident illumination. Further, he showed (**2**) that if one of the fields is more brightly illuminated than the other, the phase length of the brighter is lengthened, although there is no change in the rate of fluctuation. Breese also reported that if the interval between observations is great enough, there is no practice effect. Devries and Washburn (**3**) have corroborated the findings concerning the influence of illumination.

Washburn (**7**) has advanced the hypothesis that rivalry between the visual fields in normal vision is necessary for the perception of depth. She was able to demonstrate that rivalry could be observed, without the use of a stereoscope, while viewing a solid object (**8**). Essentially, this means

[1] An extended analysis of the results of this investigation and the present revision were performed under Contract N5-ori-166, Task Order I, between the Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-127, Project Designation No. NR-784-001, under that contract.

[2] This study is a modification of a thesis presented in partial fulfillment of the M.A. requirements in the Ohio State University. The author wishes to express his appreciation to Dr. Samuel Renshaw under whose direction the original study was performed.

that binocular rivalry is another cue for binocular depth perception.

These experimental findings may be summarized as follows: (1) when two fields which are viewed stereoscopically, and which differ sufficiently so that a unified response cannot be made to both, rivalry occurs; (2) the rate of fluctuation is increased by increasing the incident illumination while a brightness difference between the two fields merely increases the phase length of the brighter; (3) gross eye movements are not a necessary condition for rivalry to occur; (4) rivalry can be observed in the free vision of a solid object.

This study is concerned with testing the hypothesis that the rate of alternation in binocular rivalry is a function of the "strength" of the alternating figures. Koffka (4) defines figure strength in terms of resistance to distortion, impressiveness, internal articulation, density of energy, symmetry, etc. Brightness difference between figures and ground (contrast) and continuity of figure-contour are two characteristics of figure strength which can be easily manipulated. These will be used to test the hypothesis.

Procedure

All observations were made using thew Rensha Stereo-Disparator and were recorded on polygraph tape driven by a constant-speed motor. The Disparator is a modified stereoscope, the optical system of which consists of a pair of 5-diopter split spheres, mounted base out. The instrument automatically provides new and correct orthophoria positions for each change in the focal distance of the targets from the centers of the lenses. It permits a wide range of target positions simulating base-out and base-in prism effects at any focal distance between far (200 mm.) and near (131 mm.) points.

The experimental room was dimly illuminated, the main light source being a 10-w. frosted bulb on the Disparator. The polygraph was located in another room so that distracting noises were reduced to a minimum.

All the stereograms were drawn with Higgins Black Waterproof Ink. The left member of each

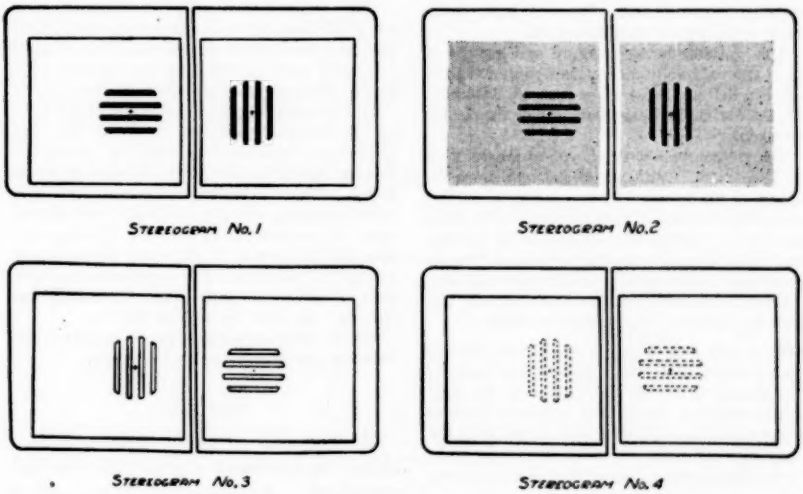

Fig. 1. Four stereograms representing different degrees of "figure strength." Stereograms 1 and 2 (S_1 and S_2) illustrate two levels of figure-ground contrast. Stereograms 3 and 4 (S_3 and S_4) illustrate two degrees of contour completeness.

TABLE I

ANALYSIS OF MEAN NUMBER OF CYCLES PER MINUTE IN THE BINOCULAR RIVALRY OF FOUR STEREOGRAMS REPRESENTING DIFFERENT DEGREES OF FIGURE STRENGTH

(Each mean is based upon ten trials)

Measure	Subjects									
	A		B		C		D		E	
Stereos.	S_1–S_3	S_2–S_4	S_1–S_3	S_2–S_4	S_1–S_3	S_2–S_4	S_1–S_3	S_2–S_4	S_1–S_3	S_2–S_4
M_1	38.15	35.30	13.60	13.90	18.20	15.30	19.15	17.45	25.90	20.90
M_2	31.65	22.50	6.15	6.45	13.05	10.35	13.85	12.70	16.00	15.75
D_m	6.50	12.80	7.45	7.45	5.15	4.95	5.30	4.75	9.80	5.15
S_{diff}	1.52	1.52	.64	.66	1.32	1.01	.74	1.27	1.51	1.14
t*	4.28	8.47	11.58	11.29	3.90	4.90	7.21	3.74	6.49	4.50

* One-tailed test. $P.01 = 3.25$; $P.001 = 4.78$; $df = 9$.

pair consisted of four parallel vertical bars, 2 mm. wide and 25 to 39 mm. long. The right member consisted of four parallel horizontal bars of similar dimensions. The bars were drawn within a 40 mm. diameter circle at the center of which was a small cross which served as a fixation point. The actual contour of the circle was not drawn.

Stereograms No. 1 (S-1), No. 3 (S-3), and No. 4 (S-4) were drawn on white drawing paper the reflectance of which was .70. Stereogram No. 2 (S-2) was drawn on grey paper (No. 27 in the Hering-Zimmerman series) which had a reflectance factor of .10. Figure 1 is a photograph of the four stereograms used.

The contrast between figure and ground in S-1 is approximately seven times greater than in S-2. In S-3 and S-4, outline bars are drawn; in S-3, the outline is continuous; in S-4, the outline is broken.

The stereograms were viewed at far point on the Disparator, a distance of 200 mm. from the centers of the lenses. At this point, all *S*s but one could comfortably fuse the targets when they were at the orthophoric separation of 95 mm. One *S* (M. F.) could only attain fusion at 80 mm. separation, indicative of a 7.5 diopter esophoria.

When *S* reported for the first time, he was given the following instructions to read:

> When you look into the eye-pieces you will see a series of vertical lines and a series of horizontal lines. In the center of the field there will be a cross. You will have three or four minutes to become acquainted with the experimental situation and to familiarize yourself with the phenomenon of binocular rivalry.
>
> After a while you will notice that the vertical lines and the horizontal lines are not seen superimposed, as a lattice, but that they appear and disappear alternately, in cycles.
>
> Fixate the cross in the center of the field and report when the cross appears between the vertical lines by pressing and releasing the key. This, together with any change from the vertical position, constitutes one cycle. A return of the vertical lines marks the beginning of the next cycle.
>
> It may be that the whole field will not alternate as a unit.
>
> Use as your criterion the region surrounding the cross.

The *S*s were seated in a chair with an armrest on which was clamped a telegraph key. They observed the rivalry phenomenon for about 1 min. with each of the four stereograms. Recorded observation began the following day.

Each session consisted of a 3-min. observation of only one stereogram of the four. The first minute was regarded as a warm-up period and was not counted, although the *S*s were not aware of this. The number of cycles in the next 2 min. were averaged, and the mean obtained constituted the score for that session. The next session was conducted in the same manner, with another stereogram. All consecutive sessions were at least 4 hr. apart. Each stereogram was observed ten times by each of five *S*s.

All *S*s were undergraduate or graduate students at the Ohio State University.

RESULTS

At the end of each session, *S* was asked to comment upon what he had observed. All reported that during most of each trial there was simultaneous rivalry in the peripheral portions of the fields. Alternation was

observed only in the region of the field about the fixation point. Occasionally, one figure would disappear completely. All *S*s mentioned the fact that when both figures were seen simultaneously, they appeared to be in different planes, one behind the other. There was no complaint of eye-strain or other kinds of fatigue.

The mean number of cycles per minute for each *S* with each stereogram is presented in Table I. In order to test the hypothesis that "strong" figures will rival faster than "weak" figures, the mean alternation rate of S-1 was compared with that of S-2, and S-3 with S-4, for each *S*. Using a *t*-test for differences between correlated means, it is possible to reject the null hypothesis in all cases well above the .01 level of confidence.

Figure 2 shows the data for a typical subject (A in Table I). Except for intertrial fluctuations, the curves maintain a fairly constant level over the ten trials. In Fig. 3 are the data for Subject E. This is the only case in which there appears to be any systematic increase in rate.

The data on S-1 and S-2 show the effect on rivalry of fields having different figure-ground brightness relationships, while S-3 and S-4 show the effect of different degrees of contour completeness. The rivalry rate of the "strong" figures, S-1 and S-3, is higher than the rate of the "weak" figures, S-2 and S-4. However, there is no uniformity among the *S*s with regard to the absolute number of cycles per minute. The range of rates is from 4 cycles per minute for Subject B on S-4 to 45 cycles per minute for Subject A on S-1.

In general, the results indicate that when viewed stereoscopically, fields

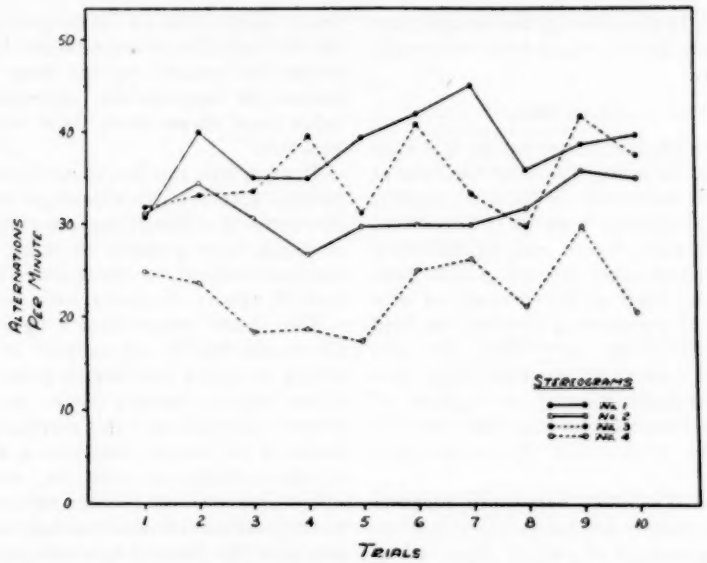

FIG. 2. Rate of alternation of rivaling figures of different "figure strength." Stereogram 1 is compared with stereogram 2, and stereogram 3 is compared with stereogram 4. (Subject A.)

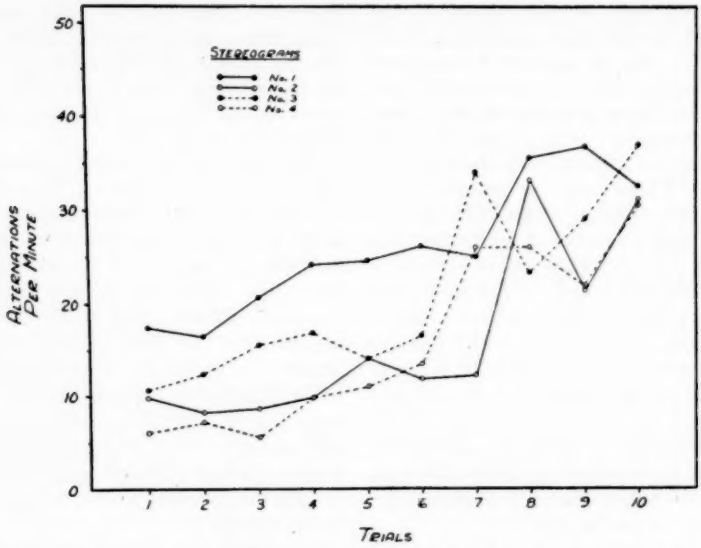

FIG. 3. Rate of alternation of rivaling figures as demonstrated by a subject with a 7.5 diopter esophoria at far point. (Subject E.)

which include strong figures alternate at a greater rate than fields with weak figures.

DISCUSSION

The present investigation was conducted to explore further the factors which influence binocular rivalry. The phenomenon seems to be related to binocular fusion, and its similarity to certain other periodic visual phenomena such as the reversal of ambiguous perspective figures, the field reorganization exemplified by the Rubin cross, and the shift from two- to three-dimensional perception of plane drawings, indicates that they all may be "explainable" by a single principle.

The Washburn hypothesis suggests a relationship between rivalry and the perception of depth in free vision. Although it may be difficult to test the Washburn hypothesis directly, it seems likely that an investigation of rivalry and of its relation to the depth perception which results from the fusion of appropriate stereoscopic fields may throw some light on the problem.

If we accept the fact that the stereograms employed in this experiment do represent different degrees of figure strength, it is possible to relate the results obtained to the Köhler-Wallach (5) theory of cortical satiation.

The theory states that a figure in the visual field is represented in the cortex by areas of different potential across which a current flows. In the course of its existence this current produces in the neural substrate a state of electrotonus, or satiation, which offers progressively greater resistance to the continued flow of the figure current or to the flow of any other current which may be induced in the same area.

The assumption is made that a figure in one half of a pair of stereograms will produce such a state of satiation, and, in the course of time, resistance will build up sufficiently to halt further flow of the current. At that point the figure will disappear and the other half of the pair will be perceived. The whole process will then be repeated.

Since it may also be assumed that a stronger figure will be represented by a stronger figural current, which will produce a state of satiation more rapidly, it may be inferred that stronger figures will alternate at a higher rate than weaker figures.

All *Ss* stated that when the two fields were seen in simultaneous rivalry, they appeared to be in different planes. This may indicate that cortical figure currents exist in three dimensions and may be associated with the emergence of the perception of depth when appropriate visual displays are combined binocularly.

The Köhler-Wallach study showed that the medium in which visual figural processes occur is anisotropic in the vertical dimension and that figural after-effects are stronger in the lower half of the field (**5**, p. 281). Comments by some *Ss* in this experiment indicate that this anisotropy affects rivalry as well. They reported that the lower half of the figure disappeared completely more often than the upper half.

The results obtained with Subject E were different from the others in that she showed increasing rates for all four stereograms during the course of the experiment. Of all the *Ss*, she alone demonstrated an anomalous space projection, an esophoria of 7.5 diopters at far point. Although this was the only outside factor measured, the relationships involved suggest further investigation of the relationship between binocular rivalry and visual skills.

Summary

An attempt was made to investigate the phenomenon of binocular rivalry in terms of figure-ground organization. The hypothesis tested was that "strong" figures will alternate more rapidly than "weak" figures when viewed stereoscopically.

Five *Ss* viewed four stereograms which exemplified two different characteristics of figure strength. Two stereograms illustrated a broken versus a continuous contour; two, greater and less contrast between figure and ground.

The data substantiate the hypothesis, and the results are discussed in terms of the Köhler-Wallach theory of cortical satiation and the Washburn theory that binocular rivalry is a necessary condition for the perception of depth.

(Manuscript received for immediate publication February 19, 1951)

References

1. Breese, B. B. On inhibition. *Psychol. Monogr.*, 1899, **3**, No. 1.
2. Breese, B. B. Binocular rivalry. *Psychol. Rev.*, 1909, **16**, 410–415.
3. Devries, A., and Washburn, Margaret F. A study of retinal rivalry in the after-image. *Amer. J. Psychol.*, 1909, **20**, 131–135.
4. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt-Brace, 1935.
5. Köhler, W., and Wallach, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
6. Peckham, R. H. Eye-movements during "retinal rivalry." *Amer. J. Psychol.*, 1936, **48**, 43–63.
7. Washburn, Margaret F. Retinal rivalry as a neglected factor in stereopsis. *Proc. nat. Acad. Sci., Wash.*, 1933, **19**, 773–777.
8. Washburn, Margaret F., and Manning, P. Retinal rivalry in free vision of a solid object. *Amer. J. Psychol.*, 1934, **46**, 632–633.

THE EFFECT OF DIFFERENTIAL NONREINFORCEMENT OF THE INCORRECT RESPONSE ON THE LEARNING OF THE CORRECT RESPONSE IN THE SIMPLE T-MAZE

BY M. RAY DENNY AND MORTON D. DUNHAM

Michigan State College

Despite extensive research on the role of reinforcement in the acquisition of an instrumental act, there has been little research on the effect of nonreinforcement of the incorrect response upon the learning of the correct response. This is true in face of the obvious and considerable use of the concept of nonreinforcement in modern theories of discrimination and trial-and-error learning. Both Spence (**5, 7**) and Hull (**2, 3**) have found it necessary to incorporate the concept of nonreinforcement into their systematic analyses and have assigned certain decremental or inhibitory properties to nonreinforcement. Spence, for example, is unable to explain transposition phenomena in stimulus-response terms, as opposed to the Gestalt patterning interpretation, without invoking inhibition through nonreward, which is levelled primarily against the incorrect response tendency.

Although it is well established that continuous[1] nonreward of a previously reinforced response will be followed by the extinction of that response, separate studies by Spence (**6**) and Denny (**1**), which were only indirectly concerned with the role of nonreinforcement of the incorrect response, lead one to question whether the nonreward of the incorrect response facilitates the learning of the correct response. In the study by Denny no difference was found in the T-maze learning of two groups of rats given differential amounts of nonreinforcement of the wrong response and an equal amount of reinforcement of the correct response. However, Denny suggests that perhaps there was no evidence of a difference because (1) secondary reinforcing cues in the delay boxes which were present just prior to the occurrence of the nonreinforcement may have offset the effect of the subsequent nonreinforcement, and (2) the nonreinforcement differential between the two groups was too small.

Assuming that the analyses of Spence and Hull are correct, the present experiment has been designed to determine the effect of nonreinforcement of the incorrect response under conditions in which secondary reinforcement is better controlled and the ratio of the amount of nonreinforcement in one group to the amount of nonreinforcement in the other group is large. When two groups that receive equal reinforcement of the correct response and unequal nonreinforcement of the incorrect response are compared, the hypothesis is that the group receiving the greater amount of nonreinforcement will learn faster, especially if secondary reinforcement of the wrong response is minimized.

[1] Unlike experimental extinction, nonreinforcement of the incorrect response in the learning of an instrumental act would not be an uninterrupted process because of the frequent occurrence of rewarded correct responses. It would also be relatively unmassed.

Experimental Procedure

Apparatus.—The apparatus used was a simple T-maze which consisted of a starting box, a combination stem and choice-point section, three interchangeable delay boxes, and two end-boxes.

The roofs of the starting box, stem, choice-point section, delay boxes and goal boxes were wood, glazed screen, translucent glass, window screen, and hardware cloth, respectively. The translucent glass arrangement at the choice point allowed *E* to follow the path of *S* through the choice point without allowing *S* to receive visual cues from the external environment. The window-screen roofs of the delay boxes were painted in such a way to allow some light to enter the box and still reduce extramaze visual cues. The sides of all alleys and boxes were 1 ft. high.

Guillotine-type doors were placed at the exit of the starting box, at the choice point, and at each end of the delay boxes. The door at the choice point was T-shaped and was designed to prevent *S* from retracing its path once a choice had been made.

The starting box, stem, and choice-point section were all painted grey, while the end-boxes were differentially constructed. The positive goal box was trapezoidal in shape, white in color, and was floored with smooth sheet metal. The negative end-box was square, black, and floored with ¼-in. hardware cloth.

Two of the delay boxes were painted grey, while the third was painted black and was floored with ¼-in. hardware cloth to correspond to the negative goal box. The window-screen roofs of the grey boxes were painted white, whereas the black delay-box roof was painted black. Immediately in front of the entrances to the delay boxes a black curtain was suspended to prevent *S* from receiving cues from the delay boxes while at the choice point.

Illumination was furnished by a 40-w. goose-necked lamp placed immediately above the stem so that it illuminated the interior of the stem and the choice point. A mirror suspended at an angle over the choice point allowed *E* to follow *S* through the choice point while he controlled the doors from his position behind the starting box.

Subjects.—The *S*s were albino rats from the rat colony of the Department of Psychology of Michigan State College. The ages of the animals at the beginning of training varied from 120 to 150 days. A total of 62 animals were used of which 17 were males and 45 females.

Preliminary training.—Beginning one week prior to the experiment, all *S*s were fed 8 gm. of Purina Dog Chow per day at the hour of the day that they were to be run in the experiment. Two days prior to the learning series each group was run in a straight alley maze consisting of the starting box, one of the grey delay boxes, and one of the two goal boxes. On each of the two days, these preliminary trials consisted of three trials to the white goal box with food and two trials to the black goal box with no food. Thus, the preliminary training consisted of a total of ten trials of which six were rewarded and four were nonrewarded.

Experimental and control conditions.—The experimental conditions, X-1 ($N = 23$) and X-4 ($N = 19$), were designed to control or eliminate secondary reinforcement of the incorrect response as much as possible. This was accomplished by using the black delay box followed by the black end-box, and the grey delay box followed by the white positive goal box. The animals in the Group X-1 received three trials per day, of which one was nonreinforced and two were reinforced. Group X-4 received six trials per day, of which four were nonreinforced and two were reinforced. Thus, each group received an equal number of reinforced trials, and an unequal number (4:1 ratio) of nonreinforced trials per day.

The control group K-4 ($N = 20$) was run under conditions designed to produce secondary reinforcement in both delay boxes. This was accomplished by using the grey delay boxes interchangeably on both sides of the choice point on both the reinforced and nonreinforced trials. It was assumed that both grey delay boxes, which were used in both the preliminary training and in the training series in conjunction with the white positive goal box, would acquire secondary reinforcing properties by virtue of being associated with the white box and also through stimulus generalization. Thus, on the nonreinforced trials in which *S* was delayed in the grey delay box prior to entry into the negative black end-box, secondary reinforcement could operate to reinforce the wrong response, and thus possibly slow down the learning of the correct response. Other than for this control, Group K-4 was treated in all respects the same as Group X-4.

The *S*s were assigned at random to the three groups.

Specific daily procedure.—On trial 1 of day 1 of the learning series, each *S* was given a free trial which was always followed by entry into a negative end-box and was thus nonreinforced. Each *S* was thereafter trained positively to the side opposite this first choice. This first response is also included in the performance measure for the first day's set of trials. In Groups X-4 and K-4 the first two trials were free-choice. On the third trial, if *S* had not made a correct response, it was forced to the correct side by blocking off the wrong alley at the choice point. After the third trial each *S* was given free choices until it had completed either four nonreinforced trials or two reinforced trials. Any remaining trials were forced. In the event *S* had made a correct response on one of the first two free-choice trials, it was continued on free trials until the completion of either four nonreinforced or two reinforced trials, and then it was forced. When the

first two trials gave correct responses, *S* was, of course, forced to the incorrect side on the remaining four trials.

Group X-1 was also given two free-choice trials each day. In this group, for *S*s that went wrong on the first trial, the forcing technique was modified for the second trial. This was necessary in order to have only one nonreinforced trial and still have a measure using the first two free trials of each day. Instead of inserting the forcing block at the choice point, the door to the delay box on the incorrect side was closed, and *S* was allowed to correct a partially wrong response and retrace its path to the correct side.[2] Corrected responses were, of course, recorded as incorrect. All other forced responses were handled in the same manner as in Groups X-4 and X-1. The criterion for having made a left or right response was contact between the rat's nose and the curtain.

All *S*s were delayed 15 sec. in the delay boxes, and 30 sec. in the negative end-box. If *S* refused to enter the end-box on nonreinforced trials, it was removed from the delay box 60 sec. after the end-box door was opened. An *S* was removed if it refused to leave the choice point in 60 sec.

Food reward consisted of one medium-sized pellet of Dickinson Dog Food, about .35 gm. At the end of each day *S*s were fed 8 gm. of Purina Dog Chow in individual cages before return to the home cage.

Each day *S*s were selected at random from the home cage so that they were not run in the same order on succeeding days of experimentation. Each *S* was trained for nine days.

Results

Learning measures.—The measures of learning were the numbers of correct responses on the initial trial and on the first two trials of each day. The initial trial was used because it provides for equalization of the number of previous food reinforcements, and because initial trials are not affected by the tendency towards spontaneous alternation. The mean number of correct responses on the first two trials of each day is, however, a more sensitive measure and gives a range of scores which is more suitable for statistical analysis. Since the relationships revealed by these two measures were found to be essentially the same, only the latter are reported here.

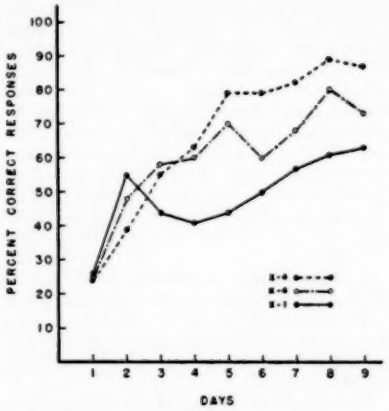

Fig. 1. Learning curves for complete groups based on the percentage of correct responses for the first two trials each day

Comparisons of learning.—The results for the three groups based on the percentage of correct responses for the first two trials each day are presented in Fig. 1. These curves show clearly that Group X-4 learns considerably faster than Group X-1. On the last day of training, Group X-4 is responding 87 per cent correctly, while Group X-1 has attained only 62 per cent cor-

TABLE I

Mean Number of Correct Responses on the First Two Trials Per Day for Days 2 to 9

Group	*N*	*M*	σ	σ_M
X-1	23	8.26	4.26	0.91
X-4	19	11.47	3.47	.82
K-4	20	10.30	2.93	0.67

[2] This procedure, if it introduces a new variable, biases the results against the hypothesis because (1) it gave Group X-1 the same number of trials as Groups X-4 and K-4 in which the choice-point cues were the same as they were on the free-choice trials, and (2) this procedure was a form of correction technique, and we know from the work of Hull and Spence (4) that the correction method gives faster learning than the noncorrection method.

rect responses. In Table I it may be seen that the over-all learning scores for the two groups, based on days 2 to 9,[3] are significantly different. The mean difference of 3.21 correct responses is significant between the .01 and .02 levels of confidence ($t = 2.63$).

Comparison of Group X-1 and Group K-4 in Fig. 1 shows that the learning curves are not as widely separated as are those of Groups X-1 and X-4. However, the mean difference between these groups for days 2 to 9 is 2.04 correct responses, which is significant between the .05 and .10 levels of confidence ($t = 2.04$). These findings agree with our expectations. As predicted, there is a clear-cut difference between Groups X-4 and X-1, in favor of X-4, and the difference between Groups K-4 and X-1 is smaller and less significant.

An examination of the learning curve for Group X-1 in Fig. 1 reveals a rapid rise in performance on day 2, followed by a sudden fall on days 3 and 4. The high score on the second day in Group X-1 may possibly be explained by the fact that all *S*s in this group were consistently reinforced to the correct side on the last two trials of the first day, whereas most *S*s of Groups X-4 and K-4 could have received secondary reinforcement of the wrong response by way of the delay box for the last two trials. The decrease in performance in Group X-1 may possibly be explained by the fact that any rewarding property originally possessed by the negative delay box extinguishes much more slowly than in Group X-4 because there is only one nonreinforcement to that side per day.

Comparisons of learning and strength of initial preference.—On the first trial of the first day of the regular training series, each *S* was given a free choice in order to establish a position preference, and each *S* was subsequently trained opposite to this preference. Inasmuch as each *S* was also given a free choice on the second trial, it was possible to estimate the relative strength of the preference. Those *S*s going left-left or right-right on the first two trials were designated as having a strong position preference, and those *S*s going left-right or right-left on the first two trials were designated as having a weak position preference. In this manner each of the three groups, X-4, X-1, and K-4, was divided into two subgroups of strong or weak preference, and a comparison of learning for each of the subgroups was made.

The results for the three subgroups with strong position preference are

[3] The data for the first day are excluded because the variable of differential nonreinforcement did not operate until after the first day of experimentation.

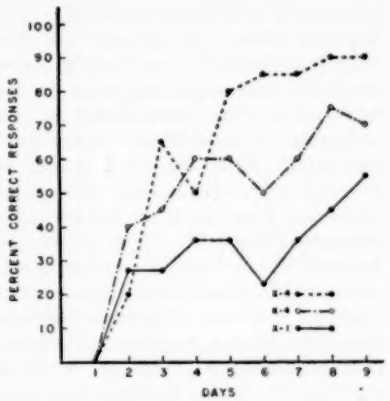

FIG. 2. Learning curves for subgroups with a strong position preference to the incorrect side as based on the first two trials each day

TABLE II

COMPARISON OF SUBGROUPS WITH A STRONG POSITION PREFERENCE IN TERMS OF THE MEAN NUMBER OF CORRECT RESPONSES ON THE FIRST TWO TRIALS PER DAY FOR DAYS 2 TO 9

Subgroup	*N*	*M*	σ	σ_M
X-1	12	5.73	3.59	1.14
X-4	10	11.30	2.86	0.95
K-4	10	9.20	2.89	0.96

given in Fig. 2 and Table II. The learning curves reveal that the X-4 subgroup learned considerably faster than the X-1 subgroup with strong preference. On the ninth day the X-4 subgroup attained 90 per cent correct responses, while the X-1 subgroup reached only 55 per cent correct responses. The K-4 subgroup falls between these two. The difference between the X-4 and X-1 subgroups is significant beyond the .01 level of confidence, and the difference between the X-1 and K-4 subgroups is significant between the .02 and .05 levels of confidence. These findings fully support those based on the entire groups. It should also be noted that the decrement in performance of Group X-1 on days 3 and 4, as observed in Fig. 1, does not occur in the subgroup composed of *S*s with a strong position preference.

One of the more obvious and significant findings in this preference analysis is that the X-1 subgroup with the strong position preference to the wrong side is definitely retarded in the learning of the correct response. Contrariwise, the X-4 subgroup with strong position preference shows a great deal of improvement and reaches a slightly higher performance level than the X-4 subgroup with weak position preference (see Figs. 2 and 3). This seems to indicate that in order for a strong, incorrect habit to be replaced by an originally weak habit within a limited number of trials, it is necessary that there be available a certain minimum number of nonreinforced trials of the wrong habit to provide for its extinction.

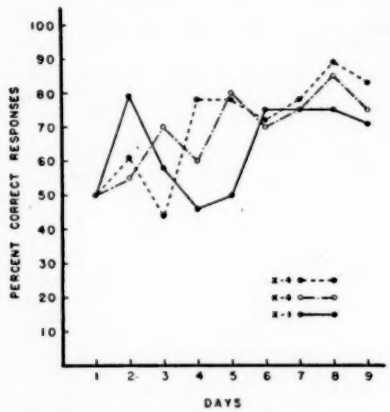

FIG. 3. Learning curves for subgroups with a weak position preference to the incorrect side as based on the first two trials each day

TABLE III

COMPARISON OF SUBGROUPS WITH A WEAK POSITION PREFERENCE IN TERMS OF THE MEAN NUMBER OF CORRECT RESPONSES FOR THE FIRST TWO TRIALS PER DAY FOR DAYS 2 TO 9

Subgroup	*N*	*M*	σ	σ_M
X-1	11	10.58	3.41	1.03
X-4	9	11.67	3.96	1.40
K-4	10	11.40	2.11	0.70

The curves for the three subgroups with weak position preference based on the percentage of correct responses on the first two trials are presented in Fig. 3 and show no clear-cut differences as a function of reinforcement-nonreinforcement schedule. Also, the curves show great variability. Noted again is the large decrement early in performance in the X-1 subgroup.

The means through learning are shown for these subgroups in Table III. None of the differences in this table is significant at better than the .50 level of confidence. These results clearly indicate that the degree of position preference plays a significant role in determining the differential effects of the three training schedules in the present experiment.

The position preference subgroup data pose the question as to what is

the relationship between speed of learning or final level of performance and the strength of the initial response tendency as measured by the direction of the initial responses in a T-maze situation (see Figs. 2 and 3). In the light of the Group X-4 data it appears that training against a strong preference will not necessarily require more trials than training against a weak preference, as current learning theory would predict. More research in this direction would seem to be warranted.

Discussion

The significant difference in learning obtained between the complete groups receiving unequal amounts of nonreinforcement of the incorrect response and equal reinforcement of the correct response indicates that nonreinforcement of the incorrect response in a simple differential response situation is an important factor in the learning of the correct response. The fact that *both* groups which received four nonreinforcements per day learned significantly better than Group X-1 seems to indicate that secondary reinforcement does not appreciably counteract the effects of subsequent nonreinforcement, at least under the temporal and stimulus conditions employed in the present study. In this context it should be pointed out that in Groups X-4 and X-1, in which the delay boxes were made differential, the differences between the boxes were not striking. Both boxes were the same size and shape and possibly indiscriminably different in brightness, since one was grey and the other black, with little light entering either box. The main difference was probably in the tactual impressions from the floor, but under-the-surface floor-cues were uncontrolled. In other words some *S*s of Group X-4 may have received secondary reinforcement for wrong responses by way of stimulus generalization for a considerable number of trials. Also, for all groups, cues at the choice point were uncontrolled and could have served to reinforce wrong responses secondarily. All this means that, despite the symmetrical properties of the maze, nonreinforcement to the incorrect side is a significant variable.

It also seems probable that whatever cues are distinctive in the total maze situation may acquire the property to mediate nonreinforcement. After a number of days of training, any secondary reinforcing value (positive expectancy) possessed by the cues on the wrong side seems to extinguish. In turn, these cues seem to acquire negative expectancy or inhibitory value. This hypothesis is based on the typical recoil behavior of *S*s in Group X-4 on the trials to the wrong side late in the learning series.

Furthermore, the X-4 *S*s who were performing near the 100 per cent correct level early in the learning series showed characteristic behavior on being forced to the incorrect side. These *S*s attempted to climb out of the maze at the choice point, ran back and forth in the stem, and upon finally entering the negative goal box, immediately tried to climb out. This frustration-type behavior and the fact that the intertrial interval was 15 min. or more may be considered as evidence against interpreting the inhibitory or extinguishing properties of the nonreinforced trials as due to reactive inhibition. Rather, it is suggested that a frustration drive state produced by repetitive nonreinforcement makes *S* avoid the cues on the wrong side and also makes possible the reinforcement of a response which avoids those cues; that is, provides additional reward for a response to the correct side. Such a

state of affairs would account for the faster learning in Groups X-4 and K-4.

The present findings support the hypotheses of Hull (**2, 3**) and Spence (**5, 7**) which assign decremental or inhibitory properties to nonreinforcement, and more specifically support Spence's contention (**5**) that nonreinforcement is largely ineffective unless considerable response tendency is available to be inhibited. They are also in full agreement with the finding of Spence and Hull (**4**) that the correction technique gives faster learning than the noncorrection technique of maze training. With the correction method a rat may receive both nonreinforcement of the incorrect response and reward of the correct response on the same trial.

Our findings confirm Denny's (**1**) earlier suggestion that his failure to demonstrate the beneficial effects of nonreinforcement of the incorrect response may have resulted from the low nonreinforcement differential between his two groups. In the earlier study one group received two reinforced and four nonreinforced trials per day, and the other group received two reinforced and two nonreinforced trials per day. Therefore, the ratio was 2 to 1 as compared to the 4-to-1 ratio in the present study, and the group which had the fewer number of nonreinforced trials in the earlier study received two nonreinforcements per day as contrasted to the one nonreinforcement per day given Group X-1 in the present study. Because the learning curve for Group X-1 is abnormally depressed (see Figs. 1 and 2), it is very likely that the crucial factor in showing that nonreinforcement is an important variable was the use of only one nonreinforcement instead of two.

Another likely reason for the difference between the two studies is the possibility that fewer *S*s in the earlier study were being trained against a strong incorrect position preference. As evidence in support of this interpretation, the per cent correct responses on the first two trials of the first learning day was approximately 20 per cent lower in the present study.

It is also true that secondary reinforcement in the delay box was uncontrolled in Denny's study, but we see from the present results that this factor only partially negated the influence of nonreinforcement.

Summary and Conclusions

The experiment was designed to test the hypothesis that, with constant reinforcement of the correct response, an increase in amount of nonreinforcement of the incorrect response in the learning of a simple **T**-maze by rats results in faster learning. An attempt was made to evaluate the influence of secondary reinforcement in this relationship.

Differential response learning, in addition to being a function of the number of reinforcements, was also shown to be a function of the number of nonreinforcements of the incorrect response.

With an increase in the strength of the preference for the incorrect response, there is an increased effect of the nonreinforcement of the incorrect response in accelerating the learning of the correct response.

If secondary reinforcement of the incorrect response precedes nonreinforcement of the incorrect response, it acts as a slight deterrent to the learning of the correct response. With better control of secondary reinforcement this effect might be made more noticeable.

The effects of nonreinforcement are interpreted in terms of a frustration drive state.

(Manuscript received April 20, 1950)

References

1. Denny, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, **36**, 373–389.
2. Hull, C. L. The concept of the habit-family hierarchy and maze learning. *Psychol. Rev.*, 1934, **41**, 33–54, 134–152.
3. Hull, C. L. Simple trial and error learning. *J. comp. Psychol.*, 1939, **27**, 233–258.
4. Hull, C. L., and Spence, K. W. "Correction" vs. "non-correction" method of trial-and-error learning in rats. *J. comp. Psychol.*, 1938, **25**, 127–145.
5. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
6. Spence, K. W. Analysis of the formation of visual discrimination habits in chimpanzee. *J. comp. Psychol.*, 1937, **23**, 77–100.
7. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W.
Washington 5, D. C.

The American Psychological Association, founded in 1892 and incorporated in 1925, is the major psychological organization in the United States. With approximately 8,500 members and 2,000 affiliates, it includes most of the qualified psychologists in the country. The purpose of the APA is to advance psychology as a science, as a profession, and as a means of promoting human welfare. It attempts to further these objectives by holding annual meetings, publishing psychological journals, conducting a personnel placement service for psychologists, and working toward improved standards for psychological training and service.

In order to give recognition to the specialized interests of different psychologists, the APA includes seventeen Divisions. Any person, after becoming a member of the APA, may apply for membership in as many Divisions as he wishes. The Divisions are:

Division of General Psychology
Division on the Teaching of Psychology
Division of Experimental Psychology
Division on Evaluation and Measurement
Division on Childhood and Adolescence
Division of Personality and Social Psychology
The Society for the Psychological Study of Social Issues—a Division of the American Psychological Association
Division on Esthetics
Division of Clinical and Abnormal Psychology
Division of Consulting Psychology
Division of Industrial and Business Psychology
Division of Educational Psychology
Division of School Psychologists
Division of Counseling and Guidance
Division of Psychologists in Public Service
Division of Military Psychology
Division on Maturity and Old Age

Each division has its own officers. Each meets annually at the time and place of the APA meeting. Each has its own membership requirements, which in some cases are higher or more specialized than the requirements for election to the APA.

The annual meetings of the APA are held in September. An attempt is made to choose meeting sites in such a way that attendance will be convenient for members and their guests in different sections of the country at different times. Thus, the 1951 meeting will be held from August 31 through September 5 in Chicago, and the 1952 meeting in Washington, D. C. These meetings provide the members with an opportunity to present or hear reports of psychological research and lectures on topics of psychological interest, and to participate in formal and informal discussion groups. The research reports, lectures, and discussion groups cover varied topics of current interest within the field of psychology.

The chief governing body of the APA is the Council of Representatives. It includes representatives from each of the Divisions. A Board of Directors, composed of the four officers of the APA (President, President-elect, Recording Secretary, and Treasurer) and six Council members elected by the Council, is the administrative agent of the Council and exercises general supervision over the affairs of the Association.

The APA maintains an office in Washington which acts as a coordinating center for all APA activities. Journal subscriptions, membership applications, personnel placement requests, and general APA business procedures are handled in this office.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	3	4	5	6	$1.25	$6.25
1921	4	1	2	3	4	5	6	$1.25	$7.00
1922	5	–	–	3	4	5	6	$1.25	$5.00
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	6	$1.25	$7.00
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	1	2	3	4	5	6	$1.25	$7.00
1951	41	By subscription, $14.00 (2 vols.)						$1.25	——
	42								
		List price, Volumes 1 through 40							$260.25
		30% discount							78.08
		Net price, Volumes 1 through 40							$182.17

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

515 Massachusetts Avenue N.W. Washington 5, D. C.